

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068666 A1

- (51) International Patent Classification: C12N 15/82
- (21) International Application Number: PCT/US01/43588
- (22) International Filing Date:
16 November 2001 (16.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/249,608 17 November 2000 (17.11.2000) US
- (71) Applicant (for all designated States except US): GENENCOR INTERNATIONAL, INC. [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): DUNN-COLEMAN, Nigel [US/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). LANGDON, Timothy [GB/GB]; Troedryhiw, Lon Tylwyd, Llanfarian, Aberystwyth SY23 4UJ (GB). MORRIS, Phillip [GB/GB]; 6 Trefaenor, Comins-coch, Aberystwyth SY23 3UB, Cerdigion, Wales (GB).
- (74) Agent: BOYD, Victoria, L.; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).
- (81) Designated States (national): AE, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, EE (utility model), EE, ES, FI (utility model), FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NR, SN, TD, TG).
- Published:
— with international search report
- (48) Date of publication of this corrected version:
6 November 2003
- (15) Information about Correction:
see PCT Gazette No. 45/2003 of 6 November 2003, Section II
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

BEST AVAILABLE COPY

WO 02/068666 A1

(54) Title: MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

(57) Abstract: Described herein are methods to enhance the production of more highly fermentable carbohydrates in plants, especially forage grasses. The invention provides for transgenic plants transformed with expression vectors containing a DNA sequence encoding ferulic acid esterase I from *Aspergillus*, preferably *A. Niger*. The expression vectors may optionally comprise a DNA sequence encoding xylanase from *Trichoderma*, preferably *T. reesei*. Expression of the enzyme(s) is targeted to specific cellular compartments, in specific cellular compartments, in specific tissues and under specific environmental conditions. Uses of this invention include, but are not limited to, forage with improved digestibility for livestock, and enhanced biomass conversion.

5

10 **MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY
OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES
ENCODING CELL WALL DEGRADING ENZYMES**

15

**Inventors: Nigel Dunn-Coleman
Timothy Langdon
Phillip Morse**

CROSS-REFERENCE TO RELATED APPLICATIONS

20 Pursuant to 35 U.S.C. §119(e), the present application claims benefit of
and priority to USSN 60/249,608, entitled "MANIPULATION OF THE PHENOLIC
ACID CONTENT AND DIGESTIBILITY OF FORAGE GRASS CELL WALLS BY
TARGETED EXPRESSION OF A FERULIC ACID ESTERASE GENE", filed
November 17, 2000, by Morris et al.

FIELD OF THE INVENTION

25 This invention relates to methods to enhance to availability of fermentable
carbohydrates.

BACKGROUND OF THE INVENTION

30 The present crisis in livestock agriculture has prompted a resurgence of
interest in grass-fed animals. However, while a high-forage diet may be desirable, it
does not currently satisfy the demands of modern animal production. For the animal
to make efficient use of the forage it consumes, the energy demands of the micro-
organisms in the rumen must be met and synchronized with the availability of plant
proteins. Otherwise this lack of synchrony will lead to (a) proteins and other
35 nutrients being poorly utilized in the rumen, (b) loss of nitrogen, in urine and feces
and therefore, the environment and (c) the need to feed excessive amounts of
protein concentrates as supplements to the ruminant diet.

Cellulose and hemicellulose in grass and maize tissues could meet the

energy requirements of the ruminant or provide new feed-stocks for industrial fermentation to ethanol. This potential is not currently realized because the cell walls are lignified and the cell wall polysaccharides highly cross-linked with phenolic residues and lignin, resulting in low rates of plant cell wall digestion in comparison to rates of protein breakdown in ruminants. This is a particular problem for the most important forages in Europe, the ryegrasses *Lolium perenne* and *L. multiflorum* as well as one of the major impediments to the wider use of better adapted species, such as *Festuca arundinacea*, as a forage crop. Increasing the digestibility index of grasses has therefore been a major breeding objective for several decades but progress has been slow due to difficulties in fixing natural variation in the synthetic varieties derived from these outbreeding species (Hayward, *et al.*, TAG 70:48 (1985)).

Removing labile phenolics by chemical treatment with alkali is known to increase the biodegradability and nutritional value of low-quality feed such as cereal straw, and is employed commercially for feed upgrading. Reducing phenolic cross-linking of cell wall carbohydrates is therefore a predictable way of improving the rate of digestion and digestibility of ryegrass. However chemical modification may have other disadvantages. Therefore, genetic modification would be a preferable method of changing the cell wall chemistry of highly digestible varieties. Many in the field are pursuing this approach. An alternative, however, is to use genetic modification to reduce the levels of phenolic acids in the cell walls available for crosslinking either by directly disrupting ester bonds linking phenolics and lignins to cell wall polysaccharides or by preventing excessive ferulation of cell wall carbohydrates prior to their incorporation into the cell wall.

This invention meets this and other needs by using targeted or inducible expression of cell wall degrading enzymes in plants.

SUMMARY OF THE INVENTION

Provided herein are methods for enhancing the availability of fermentable carbohydrates. In one aspect, there is provided an expression cassette
5 comprising a DNA sequence encoding at least one cell wall degrading enzyme. The DNA sequence encoding at least one cell wall degrading enzyme may be operatively linked to a promoter sequence. The promoter may be constitutive or inducible. The expression cassette may further comprise a targeting sequence.

In one embodiment, the cell wall degrading enzyme is selected from the
10 group consisting of ferulic acid esterase, xylanase, xylosidase, cellulase, endoglucanase, and cellbiohydrolase. In a preferred embodiment cell wall degrading enzyme is derived from a fungal source. In a more preferred embodiment, the fungal ferulic acid esterase is an *Aspergillus* ferulic acid esterase, preferably *A. niger*. In another embodiment the xylanase is derived
15 from *Trichoderma*, preferably *T. reesei*.

In another aspect of the invention, there is provided a plant transformed with the expression cassette comprising a DNA sequence encoding at least one cell degrading enzyme. The plant may be selected from the group consisting of Festuca, Lolium, Avena and Zea. In a preferred embodiment the plant is a
20 forage grass. In another embodiment, the plant is maize.

Further provided herein is a method of controlling the level of phenolic acids in plant cell walls of a transgenic plant. The method, in one embodiment, comprises introducing to a plant cell an expression cassette comprising a DNA sequence encoding at least one cell wall degrading enzyme, preferably a ferulic
25 acid esterase.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various
30 changes and modifications within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates a restriction map of a DNA fragment containing the gene encoding the 38kd ferulic acid esterase.

Figures 2 A-E illustrate the complete DNA (SEQ. ID NO:__), with
5 highlighting to point out the signal sequence, intron and various restriction
endonuclease sites, and amino acid sequence (SEQ. ID. NO:__) corresponding
to the gene encoding the 38 kD ferulic acid esterase isolated from *Aspergillus
niger*.

Figure 3 illustrates the DNA sequence of the gene encoding the 38 kD
10 esterase (SEQ. ID. NO:__).

Figure 4 illustrates the construction of the Intronless ferulic acid esterase
isolated from *Aspergillus niger*.

Figure 5 illustrates that the overlapping of PCR products made with primers FAE-I5
FAE-I3 creates two possible uninterrupted reading frames – the top in the figure below is
15 functional (highlighted serine is at active site), the bottom is inactivated.

Figure 6 illustrates the possible vector constructions useful in the present
invention. Various combinations are possible. Although and FAE gene is
depicted another cell wall degrading enzyme may be used alone (i.e., instead of)
or in conjunction with the FAE gene. Amp = ampicillin resistance gene.

20 Figure 7 illustrates pCOR105.

Figure 8 illustrates a generic ALE-TER vector.

Figure 9 illustrates the KDEL-COOH ER retention sequences.

Figure 10 illustrates the FAE-LINKER-FRAMESHIFT structure and
sequence.

25 Figure 11 illustrates plant transformation cassettes.

Figure 12 is a table of the vectors used herein.

Figure 13 depicts the barley aleurain vacuolar and apoplast signal
sequence.

Figure 14 illustrates the rat slaly transferase structure and sequence.

30 Figure 15 illustrates the potato protease inhibitor II (PPI) motif structure
and sequence.

Figure 16 illustrates the targeted expression of gfp to different cell compartment. Also shown are schematics of the vectors used.

Figure 17 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under ER and APO targeting sequences.

5 Figure 18 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under Vac targeting sequence.

Figure 19 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves of different ages.

10 Figure 20 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves under Vac, ER and APO targeting sequences.

Figure 21 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under Vac targeting sequence.

15 Figure 22 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under APO and ER targeting sequence.

Figure 23 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Festuca arundinacea* plants expressing FAE under an actin promoter.

20 Figure 24 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Lolium mutflorum* plants expressing FAE under an actin promoter.

Figure 25 illustrates the rate of fermentation and cumulative gas production in *Festuca arundinacea* cells.

Figure 26 illustrates the *in vitro* fermentation of *Festuca arundinacea* cell walls from cell cultures expressing recombinant FAE1.

25 Figure 27 illustrates the Time to maximum rate digestion for *Festuca arundinacea* cells.

Figure 28 illustrates the total gas production in *Festuca arundinacea* cells.

Figure 29 illustrates the kinetics of FAE activity by ferulic acid release from cell wall under self digestion in *Festuca arundinacea* and stimulation by xylanase.

30 Figure 30 illustrates the beta-glucoronidase activity under the *Lolium* See1 senescence promoter in leaves of transgenic plants of *Lolium mutflorum*.

Figure 31 illustrates the release of monomeric and dimeric HCAs on self digestion of leaves of vacuolar targeted FAE expressing plants.

Figure 32 is a schematic of the pTP10-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

5 Figure 33 is a schematic of the pUA4-4 vector. Also shown is the 5345 bp nucleotide sequence of the vector.

Figure 34 is a schematic of the pTU4 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

10 Figure 35 is a schematic of the pTT5.14 vector. Also shown is the 5395 bp nucleotide sequence of the vector.

Figure 36 is a schematic of the pTP8-5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 37 is a schematic of the pTP5-1 vector. Also shown is the 5277 bp nucleotide sequence of the vector.

15 Figure 38 is a schematic of the pTP4a2 vector. Also shown is the 5327 bp nucleotide sequence of the vector.

Figure 39 is a schematic of the pTP3-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

20 Figure 40 is a schematic of the pTU5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 41 is a schematic of the pGT6 vector. Also shown is the 4773 bp nucleotide sequence of the vector.

Figure 42 is a schematic of the pJQ5 vector. Also shown is the 5034 bp nucleotide sequence of the vector.

25 Figure 43 is a schematic of the pJO6.1 vector. Also shown is the 4950 bp nucleotide sequence of the vector.

Figure 44 is a schematic of the pJQ4 vector. Also shown is the 4974 bp nucleotide sequence of the vector.

30 Figure 45 is a schematic of the pPQ10.1 vector. Also shown is the 5164 bp nucleotide sequence of the vector.

Figure 46 is a schematic of the pJQ3 vector. Also shown is the 4965 bp nucleotide sequence of the vector.

Figure 47 is a schematic of the pUG4 vector. Also shown is the 5295 bp nucleotide sequence of the vector.

Figure 48 is a schematic of the pUB8.11 vector. Also shown is the 5001 bp nucleotide sequence of the vector.

5 Figure 49 is a schematic of the pTP11-1 vector. Also shown is the 5387 bp nucleotide sequence of the vector.

Figure 50 illustrates the actin promoter and its corresponding nucleotide sequence.

10 Figure 51 illustrates the Aleurain-NPIR delete structure. The corresponding nucleotide sequences are also shown.

Figure 52 illustrates the SEE1 (senescence enhanced) promoter sequence.

Figure 53 illustrates the SEE1 (senescence enhanced) promoter sequence plus the vacuolar aleurain signal/NPIR sequence.

15

DETAILED DESCRIPTION OF THE INVENTION

The invention will now be described in detail by way of reference only using the following definitions and examples. All patents and publications, including all sequences disclosed within such patents and publications, referred
20 to herein are expressly incorporated by reference.

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D Ed., John Wiley and Sons, New York
25 (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are
30 described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy

orientation, respectively. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

5 The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

Definitions

10 It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense
15 including "and/or" unless the content clearly dictates otherwise.

 "Conservatively modified variants" applies to both amino acid sequences and polynucleotides. With respect to particular polynucleotides, conservatively modified variants refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not
20 encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical polynucleotides encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the
25 corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every polynucleotide herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a polynucleotide (except AUG,
30 which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a polynucleotide which encodes a polypeptide is implicit in each described

sequence. For purposes of protein expression, there are "sub-optimal codons." These are codons that are not preferred by a particular genus or species. Altering these "sub-optimal codons" to "preferred codons" is a silent mutation in that the amino acid encoded by the codons is the same but one codon is
 5 preferentially expressed by the particular genus, *e.g.*, *Triticum spp.*

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a polynucleotide, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively
 10 modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 15 1) Alanine (A), Serine (S), Threonine (T);
 - 2) Aspartic acid (D), Glutamic acid (E);
 - 3) Asparagine (N), Glutamine (Q);
 - 4) Arginine (R), Lysine (K);
 - 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
 - 20 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (*see, e.g.*, Creighton, *Proteins* (1984)).

"Pyroglutamic acid" is the cyclized internal amide of L-glutamic acid

The phrase "controlling the level of phenolic acids" refers to the manipulation of phenolic acid expression in plants, particularly plant cell walls.
 25 The manipulation can be either positive; *e.g.*, increasing the levels of phenolic acids; negative, *e.g.*, decreasing the level of phenolic acids; or neutral, *e.g.*, changing the relative amounts of specific phenolic acids in the cell walls but keeping the total amount relatively the same. The timing of manipulation can be during plant growth or after plant growth, *e.g.*, after a plant has been cut or pulled
 30 from the ground or ingested. "Plant cell walls" refers to the cell walls of any cell of the plant.

The term "derived" means that a polynucleotide or protein is related to

another polynucleotide or protein. The relations can be one of homology, *e.g.*, nucleotides and proteins from certain species are homologous to similar polynucleotides and proteins of other species; analogy, *e.g.*, proteins perform the same function and therefore are related to each other regardless of organism of origin. The relationship can be a man-made one, *e.g.*, a protein (and a polynucleotide) can be derived from another protein by mutation; or chemical manipulation (peptidomimetics). Furthermore, a protein or a polynucleotide can be derived from an organism if, in the natural state, the protein or polynucleotide is found in one organism but recombinantly produced in another.

The term "exogenous polynucleotide" refers to a polynucleotide which is introduced into the plant by any means other than a sexual cross or sexual reproduction. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous polynucleotide is referred to here as an R₁ generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are progeny of such a plant.

The term "isolated polynucleotide molecule" or "isolated protein" refers to a polynucleotide or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *FAE1* gene is separated from open reading frames which flank the gene and encode a protein other than *FAE1*.

A "*FAE1* encoding polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of an *FAE 1* gene or which encodes a *FAE1* polypeptide. *FAE1* polynucleotides can also be identified by their ability to hybridize under low stringency conditions (see below) to nucleic acid probes having a sequence of 8 to 300 bases, preferably a sequence of 80 to 100 bases

in the sequence disclosed in WO 98/14594.

The term "nucleic acid encoding," "nucleic acid sequence encoding" or "polynucleotide encoding" refers to a polynucleotide which directs the expression of a specific protein or peptide. The polynucleotides include both the DNA strand
5 sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The polynucleotides include both full length polynucleotides as well as shorter sequences derived from the full length sequences. It is understood that a particular polynucleotide includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in
10 a specific host cell. The polynucleotide includes both the sense and antisense strands as either individual single strands or in the duplex form.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

15 The term "plasmid" refers to a circular double stranded DNA molecule which comprises the coding sequence of interest, regulatory elements, a selection marker and optionally an amplification marker. A plasmid can transform prokaryotic cells or transfect eukaryotic cells. An "expression cassette" means a portion of a plasmid (or the entire plasmid) containing the regulatory
20 elements desired for transcription, translation and/or expression and the coding region of a polynucleotide. A plasmid may contain one or more expression cassettes. If multiple expression cassettes are introduced into a plant, they can be introduced simultaneously or at different times. If simultaneous introduction is desired, the expression cassettes can be on one plasmid or more. Typically, an
25 expression cassette comprises a promoter, poly A+ tail, and signal sequences that target the expressed polypeptide to a specific region of a cell or to be secreted, if desired. Examples of signal sequences that "target expression" of ferulic acid esterase include sequences located upstream of the FAE coding sequence. The polynucleotide that encodes the signal sequence is found
30 preferably within the 100 nucleotides "upstream" (in the 5' direction) from the initiation codon (AUG). More preferably, the polynucleotide that encodes the signal sequence is found within the 50 nucleotides upstream from the initiation

codon. Many different cellular organelles are targeted by the signal sequences used in this invention. The organelles include, but are not limited to, vacuoles, Golgi apparatus, endoplasmic reticula, and apoplasts. In addition to upstream signal sequences, the expression cassette of this invention may include a
5 polynucleotide that encodes a signal sequence at the 3' end. These signal sequences include, but are not limited to stop codons and the KDEL sequence. In addition to KDEL, other similar sequences are contemplated by this invention, including but not limited to RDEL. In addition to a KDEL sequence, a signal sequence can include a linker to a KDEL sequence. A linker is an extension of
10 the reading frame of the encoding polynucleotide to the signal sequence. Preferably, the polynucleotide encoding the signal sequence is directly downstream from the coding sequence, more preferably less than 100 base pairs from the stop codon, more preferably less than 20 base pairs from the stop codon.

15 The term "polynucleotide," "polynucleotide" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses polynucleotides containing known analogues of natural nucleotides which have similar binding properties as the reference polynucleotide
20 and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular *FAE1* polynucleotide of this invention also implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions
25 may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.*, 1992; Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term polynucleotide is used interchangeably with
30 gene, cDNA, and mRNA encoded by a gene.

The term "polypeptide," "peptide," and "protein" are used interchangeably and refer to amino acids connected by peptide bonds. Polypeptides can be

entire proteins or portions thereof. For Example, a FAE1 polypeptide may refer to the entire FAE1 protein or fragments of the FAE1 protein. A "ferulic acid esterase with an altered glycosylation site" refers to a FAE protein wherein a mutation has changed the glycosylation pattern of the protein. Mutations that effect such changes are well known in the art and include, but are not limited to, amino acid substitutions, and mutations in the proteins of the Golgi apparatus and endoplasmic reticulum that effect glycosylation of proteins.

The term "promoter" refers to a polynucleotide that directs expression of a coding sequence. A promoter can be constitutive, *i.e.*, relatively independent of the stage of differentiation of the cell in which it is contained or it can be inducible, *i.e.*, induced by specific environmental factors, such as the length of the day, the temperature, *etc.* or a promoter can be tissue-specific, *i.e.*, directing the expression of the coding sequence in cells of a certain tissue type. A "senescence" promoter is an inducible promoter that causes transcription to be initiated upon a certain event relating to age of the organism. A "heat shock promoter" is an inducible promoter that causes transcription to be initiated upon a change in temperature. An example of a heat shock protein promoter is the Soybean Gmhsp promoter. In addition to these inducible promoters, one of skill will realize that other inducible promoters can be used. For example, a wound induced promoter, like LAP. See, US Patent No. 5,962,670.

The term "purified" denotes that a polynucleotide or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the polynucleotide or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The term "specifically hybridizes" refers to a nucleic acid probe that hybridizes, duplexes or binds to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and

annealing conditions, see, for example, Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) ("Sambrook") or CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

5 The term "stringent conditions" in the context of polynucleotide hybridization experiments such as Southern and northern hybridizations refers to sequence dependent, binding and washing environments. An extensive guide to the hybridization of polynucleotides is found in Tijssen (1993) LABORATORY
10 TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY—HYBRIDIZATION WITH NUCLEIC ACID PROBES part I chapter 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower
15 than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a
20 particular probe. An example of stringent hybridization conditions for hybridization of complementary polynucleotides which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin
25 with 1 mg of heparin at between 40 and 50°C, preferably 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at from 70 to 80°C with 72°C being preferable for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at
30 about 60 to 70°C, preferably 65°C for 15 minutes (see, Sambrook, *supra* for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 40 to 50°C, preferably 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6x SSC at 35 to 45°C, with 40°C
being preferable, for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Polynucleotides which do

not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a polynucleotide is created using the maximum codon degeneracy permitted by the genetic code.

5 The term "transgenic plant" refers to a plant into which exogenous polynucleotides have been introduced and their progeny. Typically, cells of a plant are transformed with the exogenous polynucleotide and a transgenic plant is regenerated from the transformed cells. The regenerated plant is then bred to produce a strain of transgenic plants.

10 "Xylanase" (EC 3.2.1.8) refers to a well described class of glycosyl hydrolases that hydrolyze xylan. Commercial applications of xylanase include the degradation and bleaching of wood pulp for paper making. Xylanase can also be added to animal feed to improve the digestibility of plant matter. Typically, commercial xylanase is derived from fungi. A preferred xylanase is derived from
15 *Trichoderma*.

Preferred Embodiments

Plant cell walls contain a range of alkali-labile ester-linked phenolic acids. In particular, grass cell walls are characterized by the presence of large amounts of esterified ferulic and *p*-coumaric acids (mainly in their *E* configurations), linked
20 to arabinoxylans at the C5 of arabinose. These are released as ferulated oligosaccharides (FAX and PAX) by cellulase treatment but *in vivo* provide a substrate for peroxidase-catalyzed cross-linking of cell wall polysaccharides and lignin. The high levels of these phenolic acids and their dimers have a dramatic influence on the mechanical properties, digestibility and rates of digestion of
25 grasses by ruminants.

Previous work has shown that ferulic acid is the predominant *p*-hydroxycinnamic acid esterified to grass polysaccharide but until recently the only ferulic acid dehydrolimer to have been isolated was 5,5'-diferulic acid. Recently new dehydrodiferulate dimers and cyclobutane-type dimer mixtures have been
30 isolated from plant cell walls (Waldron, *et al.*, *Phytochemical Analysis* 7:305 (1998)). As can be seen in Figure 1, these mixtures are present in large amounts in grass cells. Ether linked ferulic acid-coniferyl alcohol dimers, have also been isolated from cell walls (Jacquet, *et al.*, *Polyphenol Comm. Bordeaux*

pp451 (1996)) establishing for the first time that ferulate esters are oxidatively co-polymerized with lignin precursors which may anchor lignins to cell wall polysaccharides. The yield of these dimers in grass cells indicates that phenolic dehydrodimer cross-linking of cell wall polysaccharides is much more extensive
5 than was previously thought.

An enzyme system has been reported from parsley endomembranes that catalyses the ferulation of endogenous polysaccharide acceptors from feruloyl CoA, pointing to the ER/golgi as the site of polysaccharide esterification and the CoA ester as the physiological co-substrate (Meyer, *et al.*, *FEBS Lett.* 290:209
10 (1991)). Further evidence for this has been found in water-soluble extracellular polysaccharides excreted in large amounts into the medium by grass cell cultures. This material is highly esterified with ferulic and p-coumaric acid at levels similar to the cell walls of the cultured cells.

Feruloyl esterase activity has been detected in several fungal species
15 including, anaerobic gut fungi, yeasts, actinomycetes, and a few fiber-degrading ruminal bacteria, which enables them to de-esterify arabinoxylans and pectins.

Two ferulic acid esterases (FAE), distinguished on the basis of molecular weight and substrate specificity, have been isolated from *Aspergillus niger* and have been shown to quantitatively hydrolyze ferulic acid and release
20 dehydrodiferulate dimers from plant cell walls. Furthermore, FAE has been observed to act synergistically with xylanase to release ferulic acid from plant cell walls at a higher rate. Recently, a ferulic acid esterase (FAE) gene has been cloned from *Aspergillus niger* (Michelson, *et.al.* European Patent Application No. 9510370.1). The inventors have found the recombinant enzyme releases ferulic
25 acid and diferulate dimers from grass cell walls in a concentration dependent manner and that this enzyme is stable at 30°C pH 5.0 in the presence of substrate and has a half life of 61 h at 30°C in the presence of vacuolar extracts (pH 4.6) of grass cells. This gene was, therefore, a candidate for targeted and inducible expression of FAE in grasses (*e.g.*, *Lolium multiflorum*).

30 The present invention provides for methods of changing the cell wall structure of transgenic plants and therefore, making them more digestible. The method comprises introducing a ferulic acid esterase coding sequence into the

cells of a plant. Operably linked to the coding sequence is a promoter that can be either constitutive or inducible and signal sequences that serve to target expression of the coding sequence in the desired organelle in the desired cell of the plant. The signal sequences can be either or both N terminal or C terminal sequences.

Optionally, a second and/or third coding sequence is introduced into the plant. It is preferred that a fungal xylanase coding sequence be coexpressed with the FAE coding sequence.

This invention also provides for transgenic plants which contain FAE1 coding sequences, leading to more digestible grasses.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. Basic texts disclosing the general methods of use in this invention include Sambrook, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, 2ND ED. (1989); Kriegler, GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL (1990); and Ausubel *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1994)).

A. Isolation of Polynucleotides

The isolation of the polynucleotides, e.g., FAE1 and xylanase of the invention may be accomplished by a number of techniques. See, for example, copending US application 08/952,445 which describes the isolation of a FAE from *Aspergillus niger*; and copending US application 09/658,772 which describes the isolation of a xylanase from *T. reesei*.

For instance, oligonucleotide probes based on the sequences cited here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g., using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be

packaged into the appropriate vector. To prepare a library of cDNA from a specific cell culture, *e.g.*, *Aspergillus niger*, mRNA is isolated from the culture and a cDNA library containing the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based
5 upon the sequence of a known polynucleotide such as the polynucleotides cited here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. In addition to probes derived from known polynucleotides, degenerate probes may be used. Techniques for making and using degenerate probes are well known in the art
10 and can be found in Sambrook and Ausubel.

Alternatively, the polynucleotides of interest can be amplified from polynucleotide samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries
15 or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone polynucleotides that code for proteins to be expressed, to make polynucleotides to use as probes for detecting the presence of the desired mRNA in samples, for polynucleotide sequencing, or for other purposes.

Appropriate primers and probes for identifying ferulic acid esterase-specific genes, as well as xylanase sequences, from fungi and plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS,
20 (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990). Reaction components are typically: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μ M dATP, 200 μ M dCTP, 200 μ M dGTP, 200 μ M dTTP, 0.4 μ M primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec, followed by 72°C for 5 min.
25

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, *e.g.*, Carruthers, *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams, *et al.*, *J. Am. Chem.*
30

Soc. 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

5 Suitable sources for the ferulic acid esterase used in this invention include but are not limited to, *Neurospora crassa*, *Aspergillus spp.* and specifically, *Aspergillus niger*. The xylanase used in this invention can be derived from any suitable source including, but not limited to, *Trichoderma reesei* and *Aspergillus spp.*

10

B. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of plant species are well known and described in the technical and scientific literature. See, for example, Weising, *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length FAE1 protein, will preferably be combined with transcriptional and translational initiation and targeting regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant under the desired conditions.

Promoters can be identified by analyzing the 5' sequences of a desired gene. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. Messing, *et al.*, in GENETIC ENGINEERING IN PLANTS, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, *et al.*, *Plant Cell* 1:855-866 (1989); Bustos, *et al.*, *Plant Cell* 1:839-854 (1989); Green, *et al.*, *EMBO J.* 7:4035-4044 (1988); Meler, *et al.*, *Plant Cell* 3:309-316
5 (1991); and Zhang, *et al.*, *Plant Physiology* 110:1069-1079 (1996)).

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions
10 and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the actin and ubiquitin promoters and other transcription initiation regions from various plant genes known to those of skill. A particularly preferred constitutive
15 promoter is the rice actin promoter (see, McElroy, *Plant Cell*, 2:163 (1990)).

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under
20 developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots or seeds.

In one aspect of the instant invention, expression of FAE occurs after the the plant has been cut, removed from the ground or ingested. Thus an appropriate promoter would be a senescence promoter. For example, *BFN1* has
25 recently been shown to be a nuclease expressed in senescing leaves, Perez-Amador, *et al.*, *Plant Physiol.* 122:169 (2000). Similarly, SAG12, a cysteine protease is also found in senescing leaves (Noh & Amasino, *Plant Mol. Biol.* 41:181 (1999). In a preferred embodiment, the promoter from the *gem* gene of *Festuca pratensis* is used to direct expression of FAE in senescing leaves.

30 In another aspect, the FAE would be expressed upon ingestion by a foraging animal. Exemplary promoters for this aspect would include Soybean Gmhsp 17.5 promoter and the leucine aminopeptidase (LAP) promoter. The

GMhsp promoter is from a heat shock protein gene and initiates expression if the temperature of the environment is increased. In the laboratory, an increase of 15°C for 2 hours is the preferred heat shock. However, in non-laboratory conditions suitable increases in temperature will occur in silos and in the rumen of animals that have ingested the plants of this invention. The LAP promoter initiates the expression of the FAE gene upon wounding of the plant. Such wounding would occur after cutting the plant or after mastication by a foraging animal. Tissue specific promoters that could be used in this invention include promoters of genes that are differentially expressed in the leaves of grasses. An example of a leaf specific promoter is the *rbcS* promoter of tomato (*Proc. Nat'l Acad. Sci. USA* 84:7104 (1987)). This promoter normally regulates a gene determined to be important in photosynthesis.

For proper polypeptide expression, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural fungal gene, from a variety of other fungal or plant genes, or from T-DNA. These sequences are well known and readily available to those of skill in the art.

In addition to a promoter and poly A+ sequences, the preferred expression vectors of this invention also will contain signal sequences. These are polynucleotides found at the 5' and/or 3' ends of the coding region and serve to target expression of the gene to specific cellular organelles. These signal sequences can be both upstream or downstream of the coding region. Some preferred examples of upstream signal sequences include the barley aleurain sequence (Rogers, *Proc. Nat'l Acad. Sci. USA* 82:6512 (1985) which targets vacuoles and the *Aspergillus* apoplast signal. This signal sequence targets expression to the apoplast.

In addition to targeting expression to specific organelles, it may be desirable to retain the expressed FAE in the Golgi or endoplasmic reticulum. The well known ER retention signal, KDEL, can be added to the 3' end of the coding polynucleotide.

The vector comprising the expression cassettes (e.g., promoters and/or coding regions) of the invention will typically comprise a marker gene which

confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to hygromycin, kanamycin, G418, bleomycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

5

C. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment or the constructs may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

See Dalton et al. (Co-transformed, diploid *Lolium perenne* (Perennial Ryegrass), *Lolium multiflorum* (Italian Ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports (1999) 18(9), 721-726) for exemplary methods for culturing and transformation of grasses.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, et al., *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm, et al., *Proc. Natl. Acad. Sci. USA* 82:5824 (1985).

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, et al., *Science* 233:496-498 (1984), and Fraley, et al., *Proc. Natl. Acad. Sci. USA* 80:4803 (1983). US Patent 5,591,616 discloses *Agrobacterium* mediated transformation techniques in monocotyledons.

Ballistic transformation techniques are described in Klein, et al., *Nature*

327:70-73 (1987). In a preferred embodiment, a particle in-flow gun (PIG) is used to transform the plant cells of this invention.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as improved digestibility. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, *et al.*, PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE, pp. 124-176, Macmillan Publishing Company, New York, 1983; and Binding, *REGENERATION OF PLANTS, PLANT PROTOPLASTS*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, *et al.*, *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

To determine the presence of or increase of FAE1 activity, an enzymatic assay can be used or an assay to measure increases and decreases in rates of fermentation. These assays are readily available in the literature and those of skill in the art can readily find them.

One of skill will recognize that other assays can be used to detect the presence or absence of FAE1. These assays include but are not limited to; immunoassays and electrophoretic detection assays (either with staining or western blotting).

The polynucleotides of the invention can be used to confer desired traits on essentially any plant. However, the main utility of this invention is in the improved digestibility of forage plants. Thus, it is envisioned the transgenic plants of this invention will include but not be limited to the following genera *Lolium*, *Festuca*, *Triticum*, *Avena*, and *Medicago*. The FAE1 genes of the invention are particularly useful in the production of transgenic plants in the genus *Lolium*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be

introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

As mentioned above, the transgenic plants of this invention can be used as a foraging crop for animals, such as cattle, sheep, goats and horses. In addition, the methods of this invention can be used to transform any plant into which FAE expression is desired. For example, it is advantageous to break down cell walls during biomass conversion or during processing of plants for foodstuffs. This invention would help to achieve this goal more effectively and inexpensively.

10

The inventive methods herein may also be used to provide additional enzymes to enhance the availability of fermentable sugars in plants. Plant carbohydrates may be subject to further modification, either exogenously or endogenously, by the action of other enzymes. Such enzymes include, but are not limited to, endoglucanases, xylosidases and/or cellbiohydrolases. These enzymes may be provided either in an expression cassette provided for herein (i.e., endogenous) or applied to the plant cell walls (i.e., exogenous) to enhance the availability of mono- and/or di-saccharides.

Plants other than grasses may find a use in the present invention. For example, corn (or maize) is specifically contemplated to be useful. The grass Festuca is similar to maize in cell wall structure and therefore provides a good model of the ability to enhance fermentable carbohydrates in corn. Other useful plants contemplated for use in the present invention are Festuca, Lolium, Zea, Avena, Sorghum, Millet (tropical cereals), Miscanthus (a grass with potential for use as a biomass energy crop), Cenchrus, Dichanthium, Brachiaria and Paspalum (apomictic tropical range grasses) and Poa (Kentucky bluegrass).

Cell walls of forage grasses makes up 30-80% of forage dry matter representing a major source of energy for ruminants, but less than 50% of this fraction is digested by the animal. Conversion of low-value biomass to sugars and ethanol is also less than optimal due to the carbohydrate unavailability of the

feedstocks, including but not limited to bagasse, rice straw, corn stover and corn fiber.

5 Ferulic and other hydroxycinnamic acids are ester linked to arabinosyl residues in arabinoxylans, and play a key role in crosslinking xylans to lignin, resulting in less degradable cell walls. Ferulic acid esterase (FAE) can release both monomeric and dimeric ferulic acid (FA) from arabinoxylans making the cell wall more susceptible to further enzymatic attack. Transgenic plants have been produced expressing an FAE gene following microprojectile bombardment of cell cultures. Measurements of the level of FAE activity from different vectors
10 targeting FAE to the vacuole, ER and apoplast under constitutive or inducible (heat shock) promoters shows that at least for constitutive expression of vacuolar targeted FAE, the activity was highest in young leaves and increased along the leaf lamina. We also show that FAE expression results in release of monomeric and dimeric FA from cell walls on cell death and this was enhanced several fold
15 by the addition of xylanase. An effect of FAE expression on the monomeric and dimeric cell wall ester linked ferulate content in comparison to control (non-transformed) plants is seen. Generally, the lower the levels of monomers and, in particular, dimers of hydroxycinnamic acids in leaves, the higher the digestibility and/or availability of complex carbohydrates for conversion.

20 Senescence is the terminal phase in leaf development and occurs without growth or morphogenesis. Therefore the metabolism/physiology of this stage of the leaf's lifespan can be targeted directly for alteration with minimal detrimental impact on early development. Senescence follows leaf maturity and is associated with the expression of specific genes. These genes and their
25 controlling elements can be exploited to manipulate development, adaptation, productivity and quality traits in crop plants. There seems to be good conservation of senescence physiology across the range of higher plant species and thus these promoters are useful in the present invention.

30 The following preparations and examples are given to enable those skilled in the art to more clearly understand and practice the present invention. They should not be considered as limiting the scope and/or spirit of the invention, but

merely as being illustrative and representative thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); kg (kilograms); μ g (micrograms); L (liters); ml (milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); ° C. (degrees Centigrade); h (hours); min (minutes); sec (seconds); msec (milliseconds); Ci (Curies) mCi (millicuries); μ Ci (microCuries); TLC (thin layer chromatography); Et (ethyl), Me (methyl).

10

Example 1

Preparation of Enzyme Encoding DNA Sequences

A genomic clone for FAE1 (see Figures 1-3) was used as the starting point for the preparation of an intronless FAE1 encoding DNA sequence. The sequence for the genomic clone is given in Figures 2 and 3. Separate fragments for both FAE exons were recovered by PCR from a 5.5kb EcoRI fragment of the genomic clone in pLITMUS28, and 'cDNA' created by overlapping PCR. See Figure 4.

Two 5' primers were used. FAE-S5 which amplifies the entire reading frame (including the Aspergillus signal), and FAE-N5 which amplifies only the mature protein (i.e. has no signal). A number of codons are optimised (underlined in primer sequences below). The overlap product may be derived from either FAE-I5 (wild type) or FAE-I3 (conserved Ser changed to Ala) primers, allowing production of enzymatically inactive protein to check toxicity. As shown in Figure 5, overlapping of PCR products made with FAE-I5 and FAE-I3 creates two possible uninterrupted reading frames. If the complement to FAE-I5 serves as the template when recombined then the encoded protein retains the serine moiety and the esterase is functional (highlighted serine is at active site). If the FAE-I3 primer serves as the template the serine is replaced with an alanine and the esterase is inactivated (highlighted alanine in bottom amino acid sequence given in Figure 5).

25
30

Where possible, codon usage has been optimised in constructed reading frames (codon choice based on published barley preferences).

FAE-I5 (SEQ ID NO: ____)

5 GGCGCCGAGGGAGTGGCCGGTACGGTCAGCGCGTAGTCC 40-mer

FAE-I3 (SEQ ID NO: ____)

CCGGCCACGCCCTCGGCGCCTCCCTGGCGGCACTC 35-mer

FAE-N5 (SEQ ID NO: ____)

CTAAAGCTTACCATGGCGGCCGCCTCCACGCAGGGCATCTCCGA 44-mer

10 **FAE-S5 (SEQ ID NO: ____)**

CTAAAGCTTAACATGAAGCAGTTCTCCGCCAA 32-mer

FAE-3 (SEQ ID NO: ____)

TCTAAGCTTGCGGCCGCGACCGGCCAGGTGCAIGCGCCGCTCGTCATCCC
50-MER

15

Example 2

Preparation of Vectors

Vectors had the general structure shown in Figure 6.

A. Plant transformation vector series

20

Initial expression vectors were based on **pCOR105** [rice actin promoter - McElroy et al. MGG 231:150-160 (1991)] (Figure 7). pCOR105 Not and SstI sites were first destroyed [cut with NotI and SstI, followed by heat inactivation and T4 DNA polymerase treatment in the presence of dNTPs] using standard
25 methods as described in Maniatis et al. or following the manufacturer's instructions for enzymes to simplify subsequent Not cassette manipulation and allow use of unique Sst site (see below).

The *nos* terminator from **pMA406** (Ainley & Key (1990) PMB 14:949-60) was amplified by PCR using primers TER5 and TER3 to generate a fragment
30 with the following sequence (SEQ ID NO: ____):

(Pst1) (Not 1)
 (AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
 AAGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT
 5 GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT
 AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTATGATTA
 GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC
 AAAGTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATA
 AGCTT CTA GATCT (where K=G or T)
 10 (HindIII) (XbaI)

A redundancy in the TER5 primer (GCGKAA) creates fragments having either a stop codon (TAA) or glutamate codon (GAA) in one reading frame. The glutamate codon is in frame with a downstream KDEL motif.

15 The fragment and modified pCOR105 vector were cut with PstI and XbaI, according to manufacturers instructions, relevant fragments gel-purified, ligated with T4 DNA ligase and transformed into *E. coli*. Resulting clones were then sequenced to establish which TER5 alternatives were present.

Initial FAE expression vectors were then constructed from these vectors
 20 by inserting FAE-S5/FAE-3 PCR products (T4 DNA polymerase 'polished' in the presence of dNTPs, purified and digested with NotI, cloned into EcoRV and NotI digested vector) or FAE-N5/FAE-3 PCR products (purified and NotI digested, cloned into NotI digested and calf intestinal alkaline phosphatase treated vector).

The initial pCOR105-*nos* terminator clones were also modified by the
 25 addition of ALE-5/ALE-3 PCR products (encoding wild-type and modified barley aleurain signal peptides, see below for details). The products were 'polished' with T4 DNA polymerase in the presence of dNTPs, purified and cut with NotI, then cloned into EcoRV and NotI digested vectors. Addition of the ALE sequences creates a series of vectors which can express a reading frame inserted at the
 30 NotI or NcoI sites as a fusion to the barley aleurain signal, with or without vacuolar targeting motif, and with or without an ER retention motif. HindIII sites flanking the translation initiation codon and transcriptional terminator allow easy

movement of transcription units between expression vectors providing different promoter sequences. (See Figure 8 depicting the generic ALE-TER vector.)

Vector sequences were confirmed by sequencing. Two artifacts were found. Firstly, the redundant codon in TER5 was found to be AAA in one clone,
5 which was subsequently used as the source of all KDEL fusions (the peptide sequence is KPLKDEL, rather than EPLKDEL as designed). See Figure 9. Secondly, an additional base is found at the site of the redundant codon in one clone, creating a frameshifted terminal peptide (ETTEG, Figure 10) which was used as a control in some constructs.

10 Exploitation of the modular arrangement of signal peptides in the above vector series allowed various combinations of FAE and targeting motifs to be created using standard molecular biology procedures (i.e., restriction digest, purification of relevant fragments and ligation as appropriate). For example, the NotI fragment containing the FAE reading frame was inserted into the NotI site of
15 the frameshifted clone described above to create vector **pTP3.1**. The native *Aspergillus* COOH-terminus was inserted into a FAE-S5/FAE-3 clone as a SphI (T4 DNA polymerase polished) – NcoI fragment from the FAE genomic clone (replacing the NotI (T4 DNA polymerase polished) – NcoI fragment), creating vector **pTP4a2**, which then encodes the entire, unmodified, *Aspergillus* FAE.
20 Replacement of the SalI/XbaI fragment of pTP3.1 with that of pTP4a2 then created **pTP11.1**, which encodes FAE with a native *Aspergillus* COOH-terminus but a barley aleurain N-terminal signal.

Briefly, other vectors made in this series were; **pTP8.5**, the FAE NotI fragment inserted into the NotI site of an ALE-frameshifted COOH-terminus
25 construct, aleurain N-terminus; **pTP5.1**, replacement of the native *Aspergillus* COOH terminus with a KDEL peptide (NotI/XbaI fragment exchange), *Aspergillus* N-terminal signal retained; **pTU4.4**, BamHI fragment of pTP11.1 replaces BamHI fragment of pTP5.1, creates FAE reading frame fused to heterologous N- and C-termini (aleurain signal and KDEL).

30 Vectors in which the aleurain vacuolar targeting motif NPIR was replaced by NPGR (found to be inactive in some plant assays) were created by replacing an EcoRV/NotI fragment with ALE PCR product which had been cut with *AccI*

(T4 DNA polymerase polished) and NotI (vectors pTT5.5 and pTT5.14, Aspergillus COOH-terminus). The BamHI fragment of pTT5.5 was used to replace that of pTP5.1 to produce pTU5, creating an FAE reading frame fused to heterologous N- and C-termini (NPGR modification of aleurain signal and KDEL).

- 5 The aleurain signal was also modified by PCR mutagenesis to remove the vacuolar targeting NPIR motif in its entirety (directed by primer ALECUT, which contains a NotI site to allow exchange of BglII/NotI fragments). NPIR deletion was created in this way in pTP11.1 (creating pUA4.4), and in pTP5.1 by exchange of BamHI fragments with pUA4.4 (creating pUG4).

- 10 Finally, PCR mutagenesis, using overlap of fragments generated by primers GLY3 and GLYB, was also used to alter a potential glycosylation site (asparagine codon changed to aspartate, as carried out for example in Chen, H.M., C. Ford & P. J. Reilly (1994) Biochem J 301 275-281 Substitution of asparagine residues in Aspergillus awamori glucoamylase by site-directed
15 mutagenesis to eliminate N-glycosylation and inactivation by deamidation; see sequence data for exact change, vector pTP10.1).

PCR primers

- 20 **TER-5 (SEQ ID NO:___)**
AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
AAGCAGATCGTTCAAACATTTG 72-MER (The KDEL stop codon is underlined.)
TER-NOT (SEQ ID NO:___)
AAGACTGCAGACCATGGCGG 20-MER
25 **TER-3 (SEQ ID NO:___)**
AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC
ALECUT (SEQ ID NO:___)
CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC
GLYB (SEQ ID NO:___)
30 GAGGGTGTATTCGGTATCGAGTTGCAGGTTCGTATC
GLY3 (SEQ ID NO:___)
CTCGATACCCATTACACCCTCACGCCTTTCTGA

35 *B. Construction of different promoter vectors*

Various promoters were used to optimize expression and to establish constitutive, heat-shock inducibility and senescence enhancement.

i. Rice actin promoter and 1st intron

Initial vectors (Figures 11 and 12) were constructed from pCOR105 which was subsequently found to contain a 5bp deletion relative to the published sequence which destroys the *AccI* site (GTAGGTAGAC, deleted bases underlined) and may affect splicing at the adjacent 3' site. The original rice actin sequence in this region (GTAGGTAG) was therefore restored using
5 oligonucleotide NCO-ACT (CTCACCATGGTAAGCTTCTACC TACAAAAAGCTCCGCA) by replacing the *Bgl*II/*Hind*III fragment with a PCR product, to produce vector **pPQ10.1**.

A rice repetitive element is present in the upstream region of the actin promoter used in pCOR105; as this may have unpredictable effects on vector
10 expression it was removed from pPQ10.1 by deletion of the *Kpn*I/*Eco*RI fragment (end-filled with T4 polymerase and ligated following digest, restoring *Eco*RI but not *Kpn*I), to produce vector pGT6. The *Hind*III fragment containing the FAE reading frame and nos terminator of pTP3.1 (see Example 2A) was then inserted
15 into pGT6 to produce construct **pJO6.3**.

ii. Soybean heat-shock promoter

A soybean heat shock promoter from a 23kD HSP was obtained from pMA406 (Ainley & Key (1990) PMB 14:949-60). This promoter when fused to β -glucuronidase (Jefferson et al 1987 EMBO J 6:3901-3907) had previously been
20 shown to be inducible by a 10°C heat-shock and show stable expression for 24-48 hours (data not shown). β -glucuronidase fusions are a sensitive and versatile fusion marker in higher plants. The construction of the co-integration HS vectors is given below.

25

iii. Senescence enhanced expression (See1) promoter from Lolium multiflorum

The promoter and signal sequence (including NPIR motif) of the LSee1 gene was amplified from *Lolium multiflorum* cv Tribune with oligonucleotides SEE-NCO and SEE-VAC, and cloned as an *Asp*718/*Not*I replacement of the promoter region of
30 vector pTP11.1. Following sequencing to screen for PCR artifacts, one of three identical clones was chosen (pUB8.11).

The See1 promoter from maize has been cloned previously and has

EMBL accession number is AX050343. See WO0070061.

The *Lolium* version of See1 was also cloned previously (Qiang Li (2000) Studies on leaf senescence and its genetic manipulation in *Lolium multiflorum* PhD Thesis University of Wales, Aberystwyth) and has been shown to be senescence inducible when used to drive both β -glucuronidase and the *Agrobacterium* *ipt* gene.

An apoplast-targeted derivative was constructed by amplifying the Potato Protease Inhibitor (PPI) motif with primers PPI-AP6 and SEE-ATG, and cloning the product as an NgoMIV/NotI fragment into pUB8.11 (NgoMIV partial digest), to produce vector pJQ5.2. This vector has both the senescence induced promoter and the apoplast target sequence with the gene to be expressed inserted downstream of the apoplast sequence.

PCR Primers

SEE-VAC (SEQ ID NO: __)

AACCATGGCGGCCGCGCGCTCGGTGACGGGCCGGAT

SEE-NCO (SEQ ID NO: __)

TTCGGTACCATGGCCAGGTATAATTATGG

SEE-ATG (SEQ ID NO: __)

CTGCGCCGGCGAGATGGMCGTGCACAAGGAG

C. Construction of targeting sequences

In order to examine whether or not the localization of the enzyme would have an effect on the phenolic acid content of the cell wall various signal sequences were utilized. The targeting sequences were added either to the N-terminus or to the C-terminus of the gene of interest.

i. N-terminal signal sequences

Six N-terminal signal sequences were utilized:

- (a) The native *Aspergillus* end of FAE, plus excretion signal [apoplast localisation]

This is from the original clone and has the peptide sequence:

MKQFSAKHVLAVVVTAGHALAASTQGI.

- (b) The mature *Aspergillus* end, with no excretion signal [cytoplasmic localisation]

Peptide sequence is MAAASTQGI (underlined motif is common to all constructs). Truncation of the signal sequence in (a) above was carried out by PCR with mutagenic primer FAE-N5.

- 5 (c) The barley aleurain signal, including intact NPIR motif [vacuole localisation]

The barley aleurain vacuolar signal sequence (See Figure 13; Swissprot database accession number P05167) was derived entirely from overlapping primers (ALE-5, ALE-3, ALE-CUT ALE-CAP-5 and ALE CAP-3). Following primer annealing at 37°C and extension with T4 DNA polymerase in the presence
 10 of dNTPs according to manufacturers instructions, PCR with flanking primers ALE-5 and ALE-3 was carried out. The product was 'polished' with T4 DNA polymerase, purified, digested with NotI and cloned into EcoRV/NotI digested pCOR105-nos terminator vector (see above). ALE-3 contains redundancies so that clones encoding NPIR or NPGR motifs may be recovered. Two versions of
 15 the signal, with and without the vacuole targeting motif, were produced, to give putative vacuolar NPIR and apoplast (NPGR) signal sequences.

PCR Primers

- ALE-5 (SEQ ID NO: ____)**
 20 GGAATTCGTAGACAAGCTTACMATGGCCACGCCGCGTCCT 41-MER
- ALE-3 (SEQ ID NO: ____)**
 TATCCATGGCGGCCGCGCGGTGCGGTGACGGGCCGGMCGGGTTGGAGTC
 GGCGAA 55-MER
- ALE-CUT (SEQ ID NO: ____)**
 25 CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC 33-mer
- ALECAP-5 (SEQ ID NO: ____)**
 GCGACGGCGACGGCGGCCGTGGCCAGCACGGCGAGCGCCAGGAGGAGG
 ACGCGG 54-MER
- ALECAP-3 (SEQ ID NO: ____)**
 30 TCGCCGTCGCCTCCTCCTCCTTCGCGGACT 33-MER

(d) The barley aleurain signal, mutated to a NPGR motif [cytoplasmic localisation]

(e) The rat sialyl transferase golgi targeting motif [golgi localisation]

- 5 A Golgi targeting vector, pJQ3.2, was made by inserting a reading frame encoding the relevant rat sialyl transferase (RST) motif (See Figure 14. RST motif shown to function in plants by Boevink P, Oparka K, Cruz SS, Martin B, Betteridge A, Hawes C, (1998) PLANT JOURNAL 15 441-447 Stacks on tracks: the plant Golgi apparatus traffics on an actin/ER network) into vector pPQ10.1, and replacing the
- 10 EcoRI/NotI promoter/signal fragment of pJO6.3 with the fragment from this vector. Briefly, the RST motif was constructed by annealing oligonucleotides RST-F1A, RST-F1B, RST-F2A and RST-F2B, and amplifying the product with RST-5AD and RST-3A. This product was cloned and sequenced. Clones were found to have a
- 15 deletion which was corrected by PCR with RST-RPT, followed by overlap-PCR and cloning of products.

PCR primers

RST-5AD (SEQ ID NO: ____)

ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA

20 **RST-F1A (SEQ ID NO: ____)**

TTCCATGATCCACACCAACCTCAAAAAGAAGTTCTCCCTCTTCAT

RST-F1B (SEQ ID NO: ____)

AGAGTGATCACGGCGAAGAGGAGGAAGACGAGGATGAAGAGGGAGAACTTCT
TTT

25 **RST-F2A (SEQ ID NO: ____)**

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCCAA
GCCAAGGA

RST-F2B (SEQ ID NO: ____)

CATTTGGAACCTCCTTGGCTTGGAGGGTG

30 **RST-3A (SEQ ID NO: ____)**

AACCATGGCGGCCCGCCATTTGGAACCTCCTTGGCT

RST-RPT (SEQ ID NO: ____)

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCC
AAGCCAAGGA

(f) otif [cytoplasmic localisation]

(g) The potato protease inhibitor II (PPI) apoplast motif [apoplast
5 localisation]

An apoplast targeting reading frame was designed to encode the relevant
potato protease inhibitor II (PPI) motif (See Figure 15) and cloned into **pJO6.3**, to
produce vector **pJQ4.9**. Briefly, the PPI motif was constructed by annealing
oligonucleotides PPI-AP1, PPI-AP2, PPI-AP3, PPI-AP4, PPI-AP5 and PPI-AP6, and
10 cloning this product as a HindIII/NotI fragment into vector pPQ10.1; the EcoRI/NotI
promoter/signal fragment of pJO6.3 was then replaced with the equivalent fragment
from the modified pPQ10.1 vector.

PCR primers

15 PPI-AP1 (SEQ ID NO: __)
GGAATTCGTAGACAAGCTTACMATGGMCGTGACACAAGGAGGT
PPI-AP2 (SEQ ID NO: __)
GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC
PPI-AP3 (SEQ ID NO: __)
20 CCTACCTCCTGATCGTSCTCGGCCTCCTCTTGCTCGT
PPI-AP4 (SEQ ID NO: __)
CCTTGCGTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG
PPI-AP5 (SEQ ID NO: __)
GTGGACGCCAAGGCCTGCACCKCGAGTGCGGCAACCTC
25 PPI-AP6 (SEQ ID NO: __)
GGAATTCGCGGCCGCCGGCAGATGCCGAAGCCGAGGTTGCCGCACT

ii. C-terminal end signal sequences

Four C-terminal signal sequences were utilized:

30 (a) Native Aspergillus end, [CTW] (vacuole and apoplast vectors)

This was derived directly from the genomic clone (see Example 1) as a
NcoI-SphI fragment (Sph end filled with T4 polymerase) which replaces the

Nco1-Not1 region of a standard actin -FAE vector (Not1 end filled with T4 DNA polymerase).

(b) Expression vector linker alone [CTW-PVAAA] (plant optimised C-terminus for vacuole, golgi and apoplast vectors)

5 CTW is the peptide sequence of the Aspergillus FAE COOH end and is here provided by oligo FAE3. In this primer the reading frame is extended to provide the additional amino acids PVAAA which are partially encoded by the Not1 site used for cloning downstream signals see c) and d) below. Some COOH amino acids /motifs may affect compartment targeting, the PVAAA
10 sequences are expected to be neutral in this respect while the native Aspergillus end may not be.

(c) Linker plus KPLKDEL [first K is primer artifact, intended to be E] {ER retention vectors)

These sequences are provided by primer TER5 introduced during PCR to
15 generate the nos terminator fragment, and identified by sequencing within a specific clone. KDEL targeting has been demonstrated in plants by Denecke et al. ((1992) EMBO J 11: 2345-2355 Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope).

(d) Linker plus ETTEG [frameshift of (c)] (loss of ER retention - vacuole
20 vectors)

These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone (see Example 2A).

The KDEL signal is for ER retention, while others provide controls. A
25 frameshift in the TER5 region [additional A] was used in subsequent constructs to destroy the ER KDEL retention signal.

The linker used in the above C-terminal targeting sequences was PVAAA.

D. Co-integration and co-transformation vectors.

30

Co-transformation vectors

A Hygromycin resistance gene driven by a CaMV345S promoter (**pRob5**) (35S-HYG-CMV in pUC18 (modified HYG, derived from pGL2) Bilang et al (1991)

Gene 100:247-50) was used for co-transformation experiments with **pTT3** and **pTP3.1**, **pJQ4.9**, **pJQ3.2**, **pJO6.3**, **pJQ5.2**, **pUB8.1** 1 vectors.

5 *Co-integration vectors*

1. *Actin promoter constructs* - **pTR2.22**, **pTR6.1**, **pTR8.1**, **pTR9.4**, **pTR7.1**, **pTT5.5** and **5.1**.

10 The CAMV35S-hyg region from **pAJEB64TCA** [a plant expression vector constructed by Andy Bettany at IGER containing CaMV-HYG from **pTRA151** (Zheng et al 1991 Plant Physiol 97:832-835) (CaMV35S-HYG-tnl terminator as clonable cassette in **pUC4**) cloned into KpnI site of **pCOR105**] was added as a HindIII fragment at the KpnI site (T4 polymerase blunt) of **pTP4a2**, in divergent orientation
15 to FAE to create **pTR2.22**. The FAE/Nos HindIII fragment of this vector was replaced as follows in co-expression vectors. From **pTP5.1** for **pTR6.1**, from **pTP10.1** to **pTR8.1**, from **pTP11.1** to **pTR9.4**. Signal sequences of FAE in **pTR2.22** were replaced as HindIII/BglII fragments in **pTR7.1** (fragment from **pT09.1**). PCR products (**ALE5/ALE-G**) was digested with **Acc1** and T4 polymerase, polished,
20 followed by **Not1** digest and cloning into **EcoRV/Not1** digested **pTR2.22** to give clones **pTT5.5** and **5.1**.

PCR primer

ALE-G

25 TATCCATGGCGGCCGCGCGGTCTGGTGACGGGCGGCCCGGGTTGGAGTC
GGCGAA

2. *Actin promoter constructs* - **pUF1**, **pUA1K3**, **pUH4**, **pUH5**, **pUH6**, **pUH7**, **pUH8**, **pUH9**.

30 The HygR gene from **pAJEB64TCA**, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of **pTP3.1**, to give **pHOX3**. For ease of cloning the downstream HindIII site was destroyed to create **pUA1K3** and replacement of the FAE/Nos terminator HindIII fragment in this vector
35 was carried out as follows. From **pTP5.1** for **pUF1**, from **pTP11.1** for **pUH4**, from **pTP8.5** for **pUH5**, from **pTT5** for **pUH6**, from **pUA4.4** for **pUH7**, from **pTU5** for **pUH8** and from **pUG4** for **pUH9**.

3. Heat-shock promoter constructs - pUH10, pUH12, pUC5.11.

5 A co-transformation vector in which FAE is expressed from the soybean heat shock promoter was made by first modifying pMA406 to remove the nos terminator (BglII linearised and gel purified, KpnI digested, T4 DNA polymerase polished in the presence of dNTPs and recircularised), and then inserting the FAE HindIII fragment from pTP11.1, creating pTT3.1, which encodes the full aleurain signal and the native
10 *Aspergillus* COOH-terminus.

Following assays of various constructs, co-integration vectors were constructed with FAE and HygR genes arranged in tandem.

The HygR gene from pAJEB-64-TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1, to give pHOX3 and subsequently excised as a HindIII/SacI fragment (partial SacI
15 digest, relevant sites found in flanking pTP3.1 sequences) which was cloned into the HindIII/SacI sites of pMA406, in tandem orientation (vector pUH1a20). FAE sequences were then cloned into the HindIII site of pUH1a20 downstream of the heat-shock promoter (HindIII fragment from pTU5 for pUH10, HindIII fragment from pTT5 for pUH12). A pTP3.1 derivative was made by cloning the
20 CaMV/HygR HindIII cassette from pAJEB-64-TCA in tandem orientation downstream of the FAE gene in pTP3.1, inactivating the middle HindIII site by partial digestion and end-filling, and excising the combined FAE/HygR cassette as a single HindIII fragment, which was inserted at the HindIII site in pMA406 to
25 produce pUC5.11.

Example 3

Transformation of Plant Cells

Eight to ten weeks old embryogenic *F. arundinacea* and *L. multiflorum*

30 suspension cultures were bombarded either with a single co-integration plasmid DNA vector containing FAE and hyg resistance genes, or with a co-transformation vector containing FAE and with plasmid pROB5 conferring hygromycin resistance (CAMV35S-hpt- nos) using a Particle Inflow Gun (PIG) (Finer et al. (1992) Development of the particle inflow gun for DNA delivery to

plant cells Plant Cell Reports 11:323-328) and 1.5-3.0 μm gold particles as in Dalton *et al* (Dalton et al. (1999) Co-transformed diploid *Lolium perenne* (*Perennial ryegrass*), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports. 5 18: 721-726) and Kuai et al (Regeneration of fertile transgenic tall fescue (*Festuca arundinacea*) plants with a stable highly expressed foreign gene. Plant Cell Tissue and Organ Culture (1999) 58:149-154). Transformants were selected with hygromycin (25 to 50mg/l) over a 10-12 week selection period at 25°C under continuous white fluorescent light ($60 \mu\text{E m}^{-2} \text{s}^{-1}$) and plants regenerated via 10 somatic embryogenesis as in Dalton *et al* 1999, *supra*. Regenerated plants were screened for FAE activity on transfer to soil and expressing plants grown to maturity in a containment growth room at 18°C under 16h fluorescent lights ($350 \mu\text{E m}^{-2} \text{s}^{-1}$). Mature plants (6-8 weeks old) were re-assayed for FAE activity and fresh tissue harvested for Southern, Northern and Western analysis, and for self 15 digestion analysis. The remaining tissue was freeze dried and powdered for cell wall structure analysis, In vitro-dry matter digestibility (IVDMD) determinations and for in-vitro gas production determinations of rates of tissue digestion.

Example 4

Targeting of Expression Product

20 To verify that the targeting sequences are effective in delivering the gene the targeting sequences were operably linked to a green fluorescent protein GFP. The vector constructs are shown in Figure 16. Cells were transformed by particle bombardment as in Example 3. Localization of the GFP could be 25 visualized under a microscope 1 day after bombardment (i.e., shooting). See Figure 16.

Example 5

FAE1 activity

30 Plants regenerated from transformed cells showed FAE activity in all plant tissues tested. Cells were transformed as above under the direction of the ER and APO targeting sequences. FAE activity in transformed *Festuca arundinacea*

leaves of different ages was elevated compared to control (untransformed) plants. See Figures 17 and 18.

Similar results were seen with *Lolium mutiflorum* leaves at different ages transformed as above under the direction of vacuolar, ER and APO targeting sequence. See Figures 19 and 20.

FAE expression under a heat shock promoter can also be induced. (Data not shown.)

Thus, we have demonstrated FAE expression in *Festuca* and *Lolium* leaves under constitutive and HS promoters with effective FAE targeting to the vac, ER and apo.

FAE assay

FAE activity was determined in soluble extracts of fresh (or frozen at -70°C) leaves or cell cultures (0.5g) with 0.1M NaAc, pH 5.0 extraction buffer. Extracts were incubated with 24mM EF (ethyl 4-hydroxy-3-methoxycinnamate) or 1% FAXX as substrate, at 28°C for 24hrs and FAE activity calculated as the amount of ferulic acid released. FAE activity was also determined by measuring the release of monomeric and dimeric ferulic acid from self-digested leaf or cell culture samples. Fresh, or frozen, leaves or cell cultures (0.5g) were ground in 0.1M NaAc, pH5.0 extraction buffer in the presence and absence of xylanase (1000U GC140/sample) without added substrate and incubated at 28°C for 72hrs. Following incubation, and centrifugation, soluble extracts were loaded onto an activated reverse phase C18 μ Nova sep-pak column (Waters), eluted with 100% MeOH and the MeOH sample analysed by HPLC.

25

Example 6

Chemical Analysis of Cell Wall Extracts

Ester bound compounds were extracted from freeze dried powdered leaves or cell cultures (50 -100mg) with NaOH (5ml of 1M) followed by incubation at 25°C for 23hrs under N₂. After centrifugation and acidification of the soluble extract with concentrated HCl, the extracted phenolics were loaded onto an activated reverse phase C18 μ Nova sep-pak column (Waters) and eluted with 100% MeOH. and the MeOH sample analysed by HPLC.

30

HPLC was carried out with methanol: 5% acetic acid either with a 35-65% MeOH gradient in 15min (FAE assay) or with a 30-70% MeOH gradient in 25 min (monomer and dimer cell wall components) at 2ml/min on a μ Nova Pak C18 8x10 RCM (Waters). Extracts were detected and quantified with a diode array detector
5 (240-400nm Waters 996PDA) monitored at 280nm for aldehydes and 340nm for hydroxycinnamic acids. .

Levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under VAC, and ER and APO targeting sequences are reduced compared to control (untransformed) plants.
10 The results can be seen in Figure 21 and 22, respectively. Thus, we show where this does not result in reduced cell wall phenolics in growing plants with vac targeting but does result in lower phenolics with ER and apo targeting. In addition,

Levels of esterified monomeric and dimeric hydroxycinnamic acids in
15 *Festuca arundinacea* plants expressing FAE are not significantly reduced when FAE is VAC, targeting (Fig 21) which is as predicted for correct vacuolar targeting, but are significantly reduced, as predicted, in some plants when FAE was ER and APO targeted, compared to control (untransformed) plants. The results can be seen in Figure 22.

20

Example 7

In vitro dry matter digestibility. (IVDMD)

The *in vitro* dry matter digestibility (IVDMD) was estimated on 1.0 g dry weight of leaf or cell culture tissue using the pepsin/cellulase method of Jones
25 and Hayward (The effect of pepsin treatment of herbage on the prediction of dry matter digestibility from solubility in fungal cellulase solutions. Journal of the Science of Food and Agriculture (1975) 26:711-718).

We show that the presence of FAE in the plants results in higher digestibility of the leaves. This may be due to internal FAE activity acting on
30 normal cell walls with vacuole located FAE and to both FAE activity and the lower cell wall crosslinking with ER and apo targeted FAE (as also found with cell cultures).

End point digestibility as determined by IVDMD were higher in leaf tissue of some transformed plants of *Festuca* expressing FAE, compared to control (untransformed) plants. Examples are shown where vacuolar, ER or apoplast targeted FAE under a constitutive actin promoter have been effective at increasing IVDMD. Similar results were obtained with in leaves of *Lolium*, but were less pronounced.

The results can be seen in Figures 23 and 24.

Example 8

In vitro gas production measurements

In each experiment, 1.0-g samples of freeze dried powdered leaf tissue or cell culture were fermented in three 165-ml capacity serum bottles according to the pressure transducer technique of Theodorou et al. (Theodorou et al. (1994) A new gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. *Animal Feed Science and Technology* 48: 185-197). Grab samples of rumen-digesta were taken at 8.00 h before the morning feeding from fistulated wethers fed grass hay, and transported to the laboratory in a pre-warmed (39°C) vacuum flask. The microbial inoculum and culture media were prepared as described by Theodorou et al. (1994). Each serum bottle received 10 ml of microbial inoculum, 85 ml of buffer and 4 ml of reducing agent.

At the end of the incubation period, (144h) the contents of each serum bottle were filtered through pre-weighed sintered glass funnels and freeze dried to constant weight. Dry matter loss was calculated as the difference between the dry weight of the sample pre- and post-incubation. Additionally, the concentration of volatile fatty acids (VFA) in the liquid fraction of the culture media at the end of the 144-h incubation period was determined by gas chromatography. A Chrompack CP 9000 chromatograph fitted with an automatic sampler (Chrompack 911) and a flame-ionisation detector, linked to a Dell PC with A1-450 integration software, was used for VFA quantification.

Gas production data were fitted to the model of France et al. (France, J., Dhanoa, M.S., Theodorou, M.K, Lister, S.J., Davies. D.R. and Isac, D. 1993. A model to interpret gas accumulation profiles associated with *in vitro* degradation of ruminant feeds. *Journal of Theoretical Biology*. 163: 99-111.) using the MLP (Ross,

G.J.S. 1987: *MLP, Maximum Likelihood Program Version 3.08*. Oxford Numerical Algorithms Group) package. The equation is in the form, $Y = A\{1 - e^{[-b(t-T) - c(\sqrt{t}-\sqrt{T})]}\}$ where Y is the cumulative gas production (ml), A is the asymptote (i.e. gas pool), T is lag time, and b (h^{-1}) and c ($\text{h}^{-0.5}$) are decay rate constants. A combined fractional rate (h^{-1}) of gas production (μ) was calculated as, $\mu = b + c/2\sqrt{t}$, where t is the incubation time (h).

It can be seen for *Festuca arundinacea* (denoted as BN in Figure 25) that cell cultures have a higher rate of digestion and cumulative gas production in the presence of FAE and that the addition of an exogenous xylanase further enhance the availability of fermentable carbohydrates. Similar results are found in FAE expressing cultures without added FAE. Fermentation rates are further increased compared with controls by the addition of exogenous FAE or xylanase as these cultures expressing FAE have a reduced cell wall phenolic composition to controls Figures 26-28.

15

Example 9

FAE & xylanase transformed plants

Addition of exogenous xylanase (GC140) greatly increased FAE mediated release of phenolics from *Festuca* and *Lolium* leaves expressing *A. niger* FAE. See Figures 29-31 which show that phenolic release from leaf cell walls is increased in all FAE expressing plants on cell death and this is stimulated by xylanase irrespective of the targeting. Therefore expression of a fungal xylanase in plant cells is tested.

The FAE expression cassette is modified to comprise a fungal xylanase gene (either *T. reesei* or *A. niger*) to yield a FAE-xylanase expression cassette. The FAE-xylanase expression cassette is used to transform plant cells in a manner similar to those described in Example 3. The transformed cells are allowed to grow and are selected on an appropriate medium. The enzymes so expressed increase the availability of fermentable carbohydrates to a greater extent than the FAE expression cassette.

30

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will

be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

What is claimed:

1. A transgenic plant comprising an expression cassette comprising a promoter operably linked to a ferulic acid esterase encoding polynucleotide.
5
2. The plant of claim 1, wherein the polynucleotide is derived from *Aspergillus niger*.
3. The plant of claim 2, wherein the polynucleotide is FAE ! from *Aspergillus niger*.
- 10 4. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with an altered glycosylation site.
5. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with a substitution so that glycosylation is altered.
- 15 6. The plant of claim 3, wherein the polynucleotide further comprises a polynucleotide that encodes CTWPVAAA at the 3' end.
7. The plant of claim 3 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.
8. The plant of claim 1, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by sexual reproduction.
- 20 9. The plant of claim 1, wherein the promoter is an inducible promoter.
10. The plant of claim 9, wherein the promoter is a senescence promoter.
11. The plant of claim 9, wherein the promoter is a heat shock
25 promoter.
12. The plant of claim 1, wherein the promoter is a constitutive promoter
13. The plant of claim 1, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the

polynucleotide.

14. The plant of claim 13, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

15. The plant of claim 14, wherein the polynucleotide is derived
5 from the signal sequence of a vacuolar targeted gene

16. The plant of claim 15, wherein the targeted gene is a barley aleurain gene.

17. The plant of claim 15, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum or apoplast
10 signal sequence.

18. The plant of claim 15, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted senescence gene.

19. The plant of claim 18, wherein the senescence gene is a *Lolium Sec1* signal sequence.

20. The plant of claim 13, wherein the polynucleotide is derived
15 from the signal sequence of a golgi targeted gene.

21. The plant of claim 20, wherein the targeted gene is a rat sialyl transferease signal sequence.

22. The plant of claim 13, wherein the polynucleotide is derived
20 from the signal sequence of an apoplast signal sequence.

23. The plant of claim 22, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

24. The plant of claim 16, wherein the polynucleotide is derived from *Solanum tuberosum*.

25. The plant of claim 13, wherein the polynucleotide sequence
25 is downstream of the C-terminus of the ferulic acid esterase polynucleotide

26. The plant of claim 25, wherein the polynucleotide sequence is a KDEL sequence.

27. The plant of claim 25, wherein the polynucleotide sequence is a stop codon.

28. The plant of claim 25, wherein the polynucleotide sequence is an extension of the ferulic acid esterase reading frame to provide a linker to
5 KDEL.

29. The plant of claim 1, further comprising introduction into the plant a second expression cassette comprising a promoter operably linked to a xylanase encoding polynucleotide.

30. The plant of claim 29, wherein the xylanase encoding
10 polynucleotide is from *Trichoderma reesei*.

31. The plant of claim 29, wherein the first and second expression cassettes are present on separate plasmids.

32. The transgenic plant of claim 1, selected from the group consisting of Festuca, Lolium, Zea and Avena.

33. The transgenic plant of claim 32, wherein the plant is a
15 Festuca plant.

34. A method of controlling the level of phenolic acids in plant cell walls of a transgenic plant, the method comprising introducing into the plant an expression cassette comprising a promoter operably linked to a ferulic acid
20 esterase encoding polynucleotide.

35. The method of claim 34, wherein the polynucleotide is derived from *Aspergillus niger*.

36. The method of claim 35, wherein the polynucleotide is a FAE
1 gene from *Aspergillus niger*.

37. The method of claim 36, wherein the polynucleotide encodes
25 the ferulic acid esterase with an altered glycosylation site.

38. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with a substitution such that glycosylation is altered.

39. The method of claim 36, wherein the polynucleotide

comprises CTWPVAAA at the 3' end.

40. The method of claim 36 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.

41. The method of claim 36, wherein the polynucleotide
5 comprises SEQ ID NO:1.

42. The method of claim 34, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by transformation of cell cultures.

43. The method of claim 42, wherein the cell cultures are
10 regenerated to plants.

44. The method of claim 34 wherein the ferulic acid esterase polynucleotide is introduced into the plant by sexual reproduction.

45. The method of claim 34, wherein the transgenic plant is a member of a genus selected from the group consisting of *Festuca*, *Lolium*,
15 *Avena* and *Zea*.

46. The method of claim 45, wherein the transgenic plant is a member of the genus *Festuca*.

47. The method of claim 46, wherein the transgenic plant is a *Festuca arundinacea*.

48. The method of claim 34, wherein the promoter is an
20 inducible promoter.

49. The method of claim 48, wherein the promoter is a senescence promoter.

50. The method of claim 48, wherein the promoter is a heat
25 shock protein promoter.

51. The method of claim 34, wherein the promoter is a constitutive promoter.

52. The method of claim 51, wherein the promoter is an actin

promoter.

53. The method of claim 34, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the polynucleotide.

5 54. The method of claim 53, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

55. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene.

10 56. The method of claim 55, wherein the targeted gene is a barley aleurain gene.

57. The method of claim 55, wherein the polynucleotide is derived from the signal sequence of a Lolium See1 signal sequence.

15 58. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum signal sequence.

59. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce an apoplast signal sequence.

20 60. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a golgi targeted gene.

61. The method of claim 60, wherein the targeted gene is a rat sialyl transferease signal sequence.

25 62. The method of claim 59, wherein the polynucleotide is derived from the signal sequence of a fungal apoplast signal sequence.

63. The method of claim 62, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

64. The method of claim 59, wherein the apoplast signal

sequence is from a potato.

65. The method of claim 53, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

5 66. The method of claim 65, wherein the polynucleotide sequence is a KDEL sequence.

67. The method of claim 65, wherein the polynucleotide sequence is a stop codon.

68. The method of claim 65, wherein the polynucleotide
10 sequence is an extension of the ferulic acid esterase reading frame to provide a linker to KDEL.

69. The method of claim 34, further comprising simultaneous introduction into the plant a second expression cassette comprising a promoter operably linked to a polynucleotide encoding a xylanase gene.

15 70. The method of claim 69, wherein the second polynucleotide is a fungal xylanase.

71. The method of claim 70, wherein the fungal xylanase is from *Trichoderma reesei*.

72. The method of claim 35, wherein the first and second
20 expression cassettes are present on separate plasmids.

73. The method of claim 1, wherein the first and second expression cassettes are present on separate plasmids.

74. A transgenic plant produced by the method of claim 34.

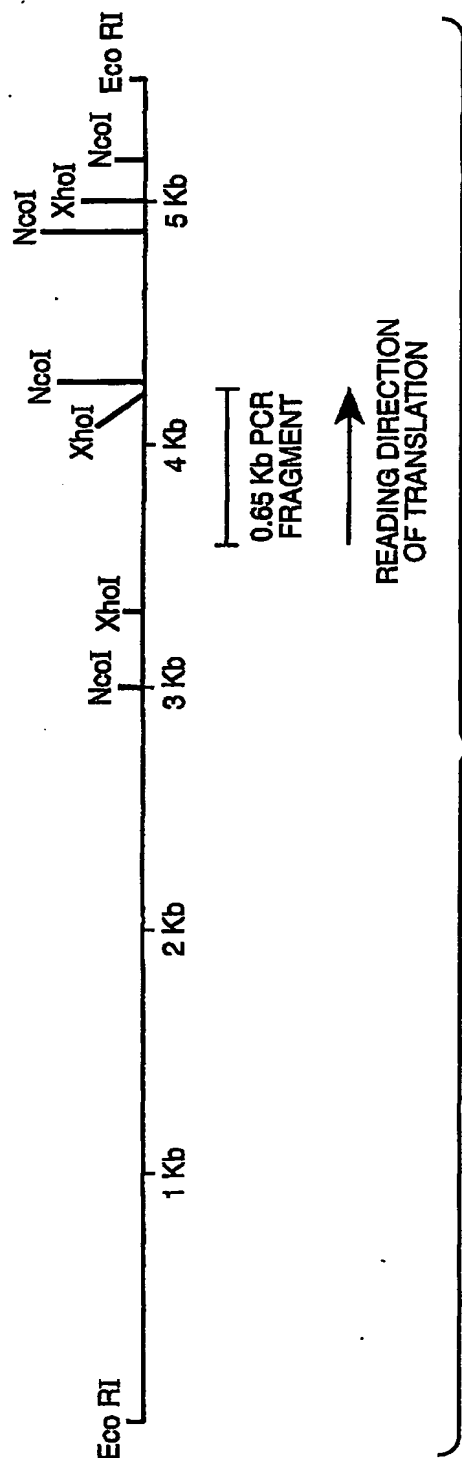


FIG. 1

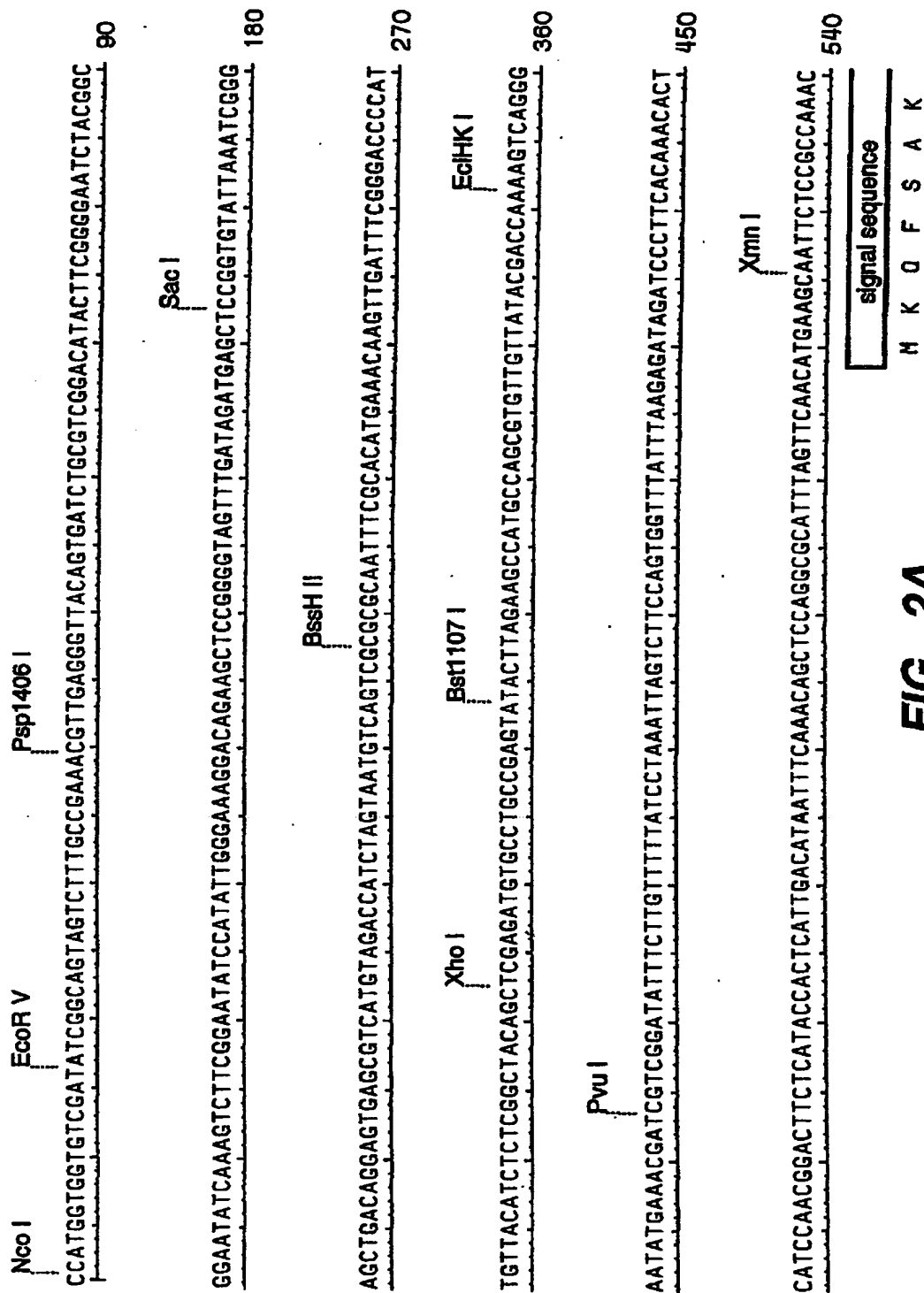


FIG. 2A

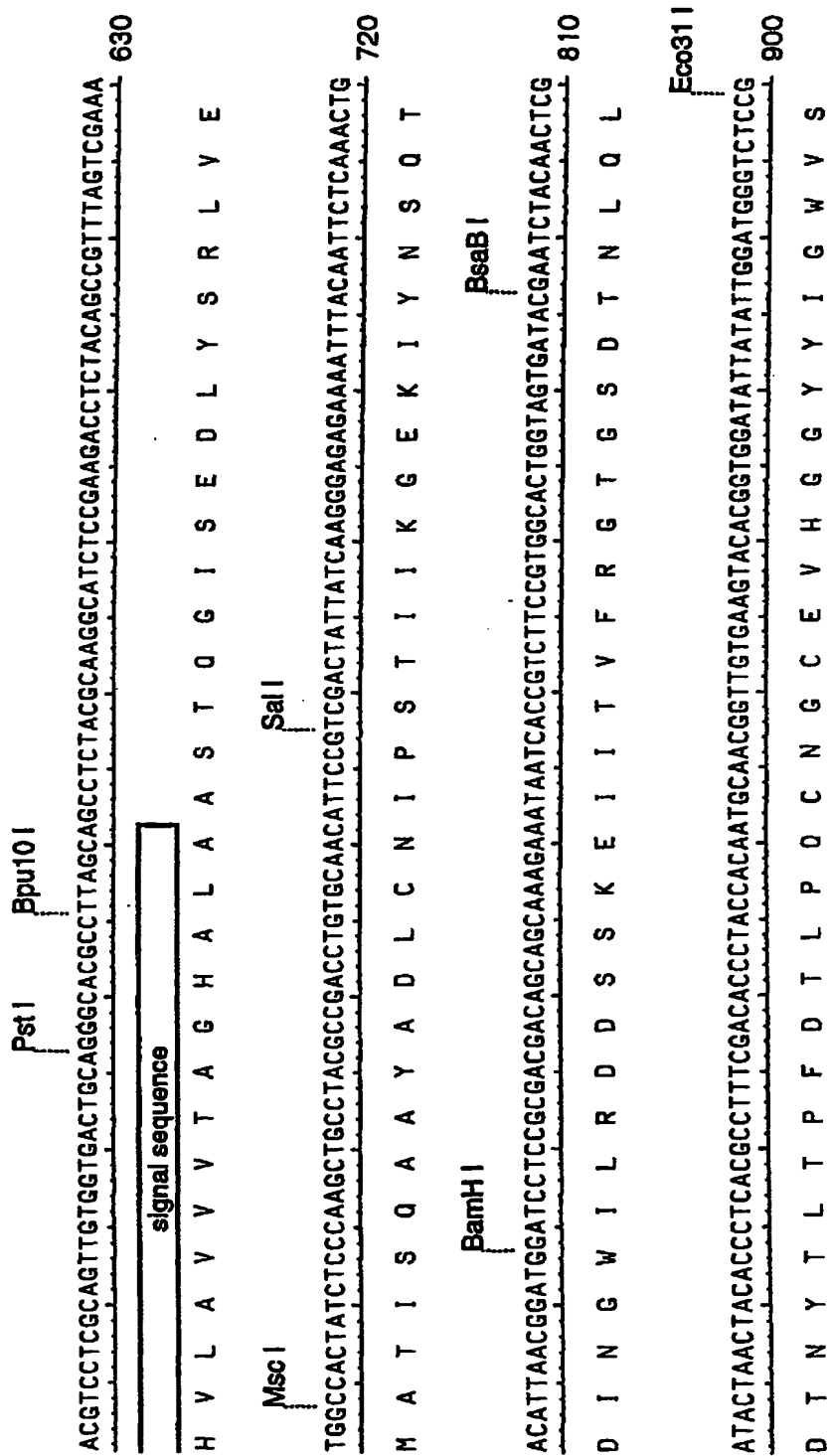


FIG. 2B

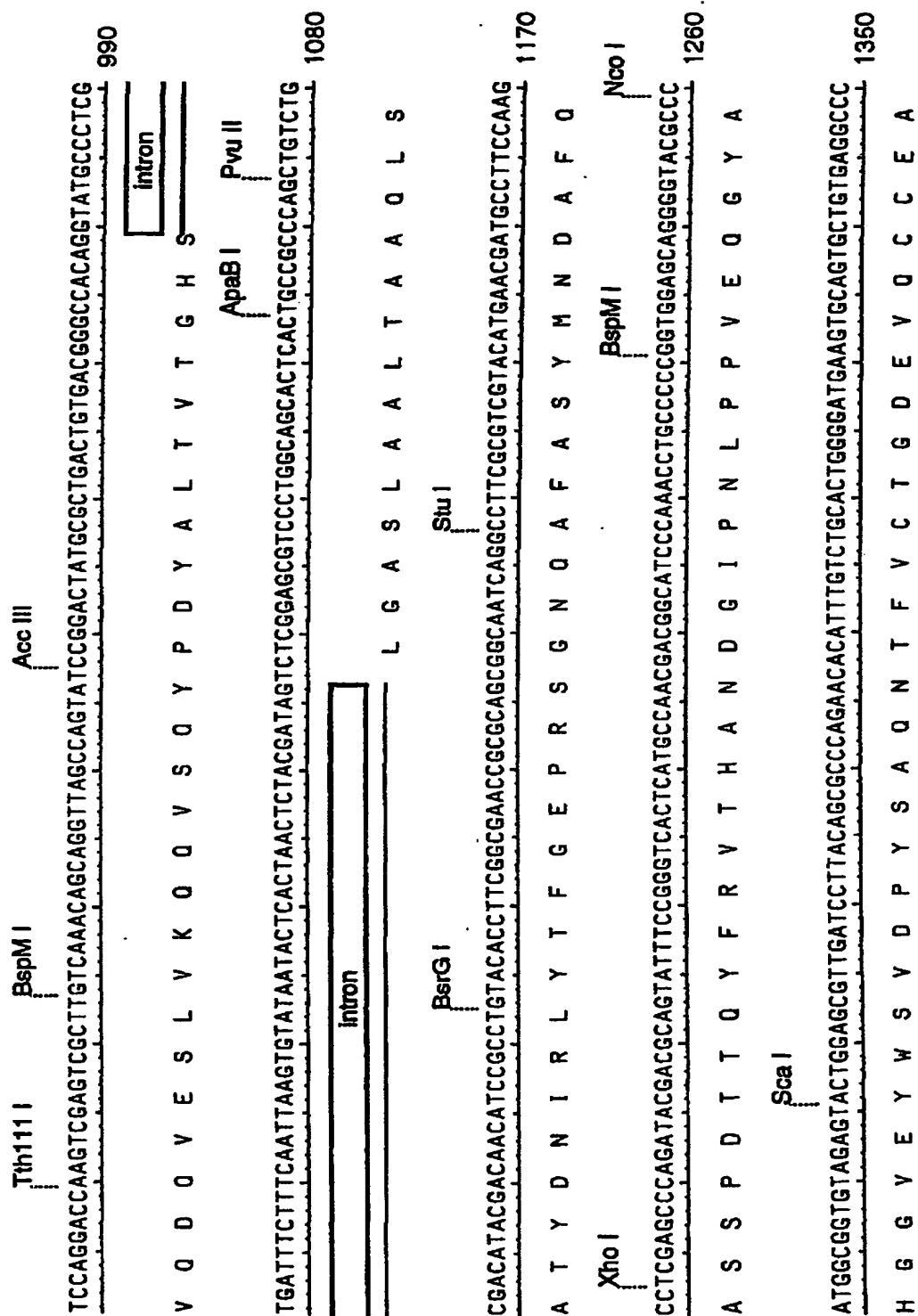


FIG.-2C

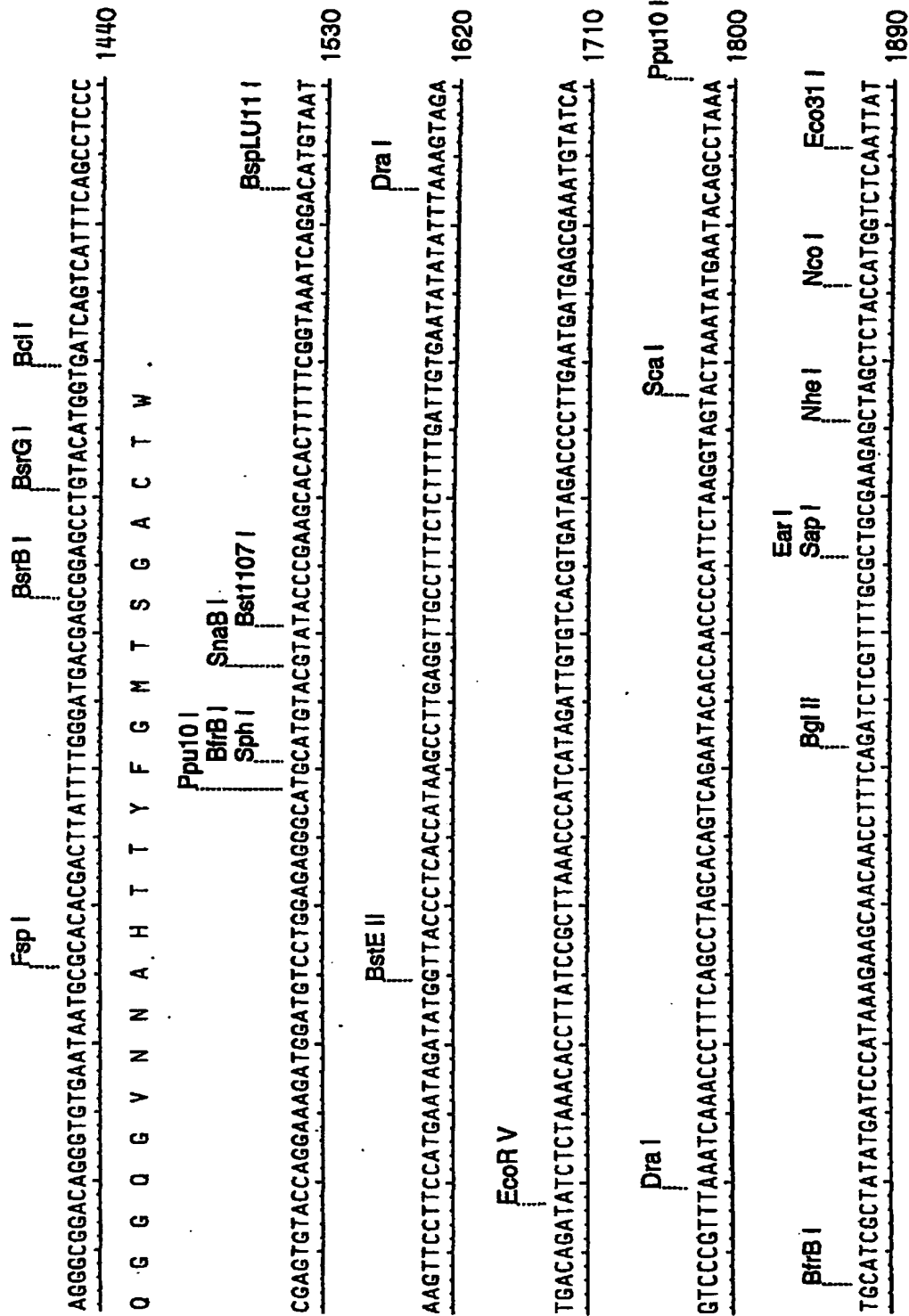


FIG._2D

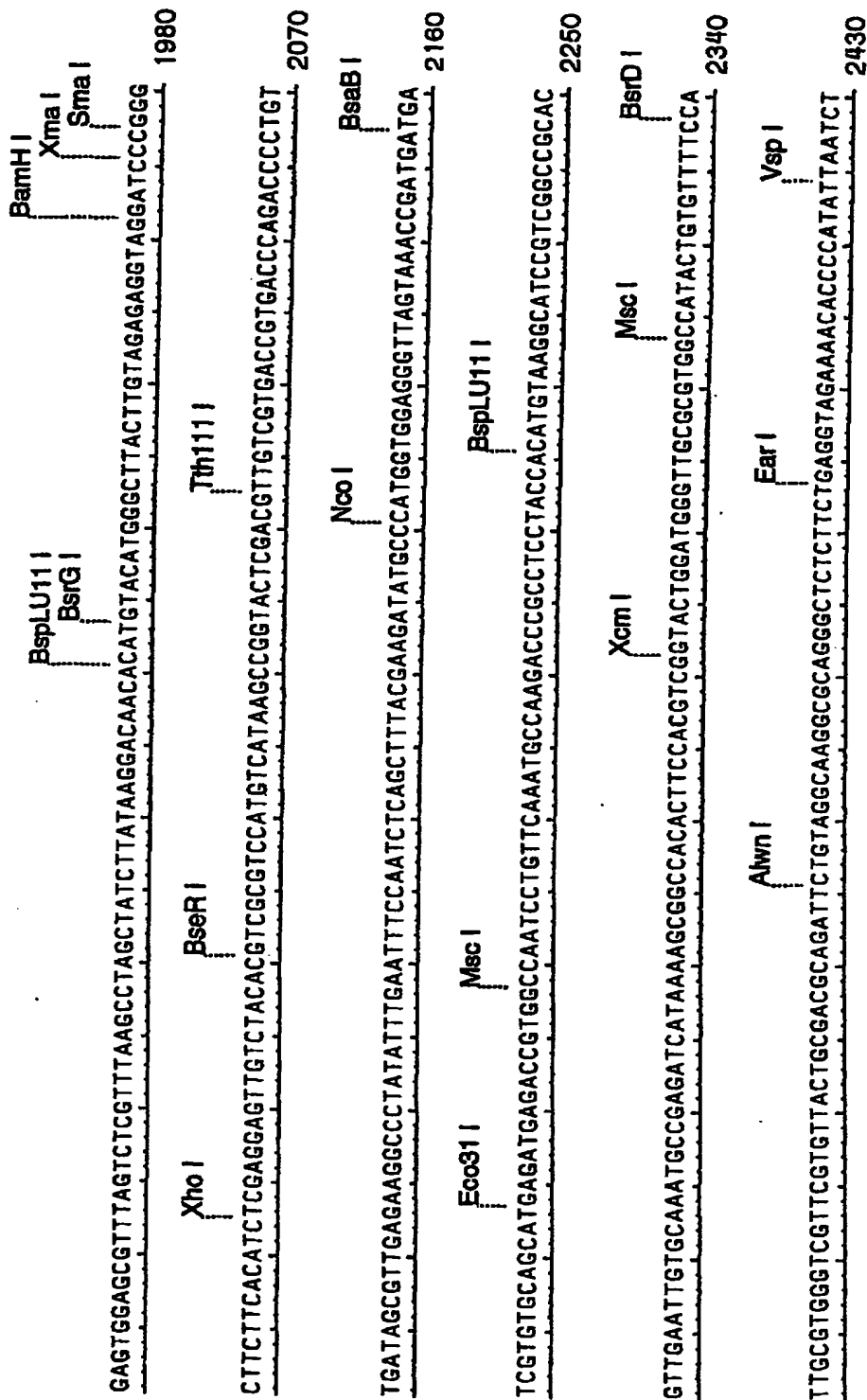


FIG. 2E

CCATGGTGGTGCATATCGGCAAGTAGTCTTTGGCCGAAACGTTGAGGGTTACAGTGATCTGCGTCGGACATACCTT
CGGGAAATCTACGGCGGAATATCAAAGTCTTCGGAATATCCATATTTGGGAAAGGACAGAAAGCTCCGGGGTAGTTT
GATAGATGAGCTCCGGTGATTTAAATCGGAGCTGACAGAGTGAGCGTCAATGTAGACCATCTAGTAATGTCAGT
CGCGCGCAATTTCCGACATGAACAAGTTGATTTCCGGGACCCCATTTGTTCATCTCTCGGCTACAGCTCGAGATG
TGCCTCGCGAGTATACTTAGAAGCCATGCCAGCGGTGTTGTTATACGACCAAAAGTCAGGGAATATGAACGATCG
TCGGATATTTCTGTTTATATCTTAATTAAGTCTTCAGTGGTTTATTTAAGAGATAGATCCCTTCACAAACACT
CATCCAACGGACTTCTCATACCACTCATTTGACATAAATTTCAAACAGCTCCAGGCGCATTTAGTTCAACATGAAGC
AATTCCTCCGCCAAACACGTCCTCGCAGTTGTGGTGACTGACAGGACCGCTTAGCAGCCCTCTACGCAAGGCATCT
CCGAAGACCTCTACAGCCGTTTAGTCGAAATGGCCACTATCTCCCAAAGCTGCCCTACGCCGACCTGTGCAACATTC
CGTCGACTATTATCAAGGAGAGAAAATTTACAATTTCTCAAACCTGACATTAACGGATGGATCCTCCGACGACA
GCAGCAAGAAAATAATCACCGTCTTCGTTGGCACTGGTAGTGATACGAATCTACAACTCGATACATACTACACCC
TCACGCCTTTCGACACCCCTACCAACATGCAACGGTTGTGAAGTACACGGTGGATATATATTGGATGGGTCTCCG
TCCAGGACCAAGTCGAGTCGCTGTGTCAAAACAGCAGGTTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC
ACAGGTATGCCCTCGTGATTCTTTCAATTAAGTGATATAATACTCACTAATCTACGATAGTCTCGGAGCGTCCC
TGGCAGCACTCACTGCCGCCAGCTGTCTGCGACATACGACAACATCCGCCCTGTACACCTTCGGCGAACCCGCGCA
GCGGCAATCAGGCCCTTCGCTCGTACATGAACGATGCCCTCCAAAGCTCGAGCCCAAGTACGACGATATTTC
GGTCACCTCATGCCAACGACGGCATCCCAAACCTGCCCGGTGGAGCAGGGGTACGCCCATGGCGGTGTAGAGT
ACTGGAGCGTTGATCCTTACAGCGGCCCAGAACACATTTGTCTGCACTGGGGATGAAGTGCAGTGTGTGAGGCC
AGGGCGGACAGGGGTGAAATAATGCGCACACGACTTATTTTGGGATGACGAGCGGAGCCCTGTACATGGTGATCAG
TCATTTTCAGCCCTCCCGAGTGTACCAAGGAAAGATGGAATGTCCTGGAGAGGCGCATGCAATGTACGTATACCCGAAAGC
ACACTTTTTCGGTAAATCAGGACATGTAAATAAGTTCTTCCATGAATAGATATGGTTACCTCAACCATAGCCCTT
GAGGTTGCCCTTCTCTTTGATTTGTGAATATATATTTAAAGTAGACAGATATCTCTAAACACCTTATCCGCT
TAAACCCATCATAGATTGTGTACCGTGATAGACCCCTTGAATGATGAGCGGAAATGTATCAGTCCCGTTTAAATCA
AACCCTTTCAGCCTAGCACAGTCAGAAATACACCAACCCCATTTCTAAGGTAGTACTAATAATAGATACAGCCATAA
TGCAATCGCTATATGATCCCATAAAGAAAGCAACACCTTTCAAGATCTCGTTTTCGCTGCGAAGAGCTAGCTCTAC
CATGGTCTCAATATAGTGGAGCGTTTAGTCTCGTTTAAAGCCTAGCTATCTTATAGGACAACACATGTACATG
GGCTTACTTGTAGAGAGGTAGGATCCCGGCTTCTTCACATCTCGAGGAGTTGTCTACACGTCGCGTCCATGTCA
TAAGCCGGTACTCGACGTTGTGTCGTGACCGTGACCCAGACCCCTGTGATAGCGTTGAGAGGCCCTATATTGAA
TTTCCAAATCTCAGCTTTACGAAAGATATGCCCATGGTGGAGGGTTAGTAAACCGATGATGATCGTGTGCAGCATGA
GATGAGACCGTGGCCAAATCCTGTTCAAATGCCAAGACCCGCTCTACCATGTAAAGGCATCCGTCGGCCGAC
GTTGAAATGTGCAAAATGCCGAGATCATAAAGCGGCCACACTTCCACGTCGGTACTGGATGGGTGCGGTGGCC
ATACTGTGTTTCCATTGCGTGGGTGCTTCGTGTTACTGCGACGCAATCTCTGTAGGCAAGGCGCAGGGCTCTCT
TCTGAGGTAGAAAACACCCCATATTAATCTGAATTC

FIG._3

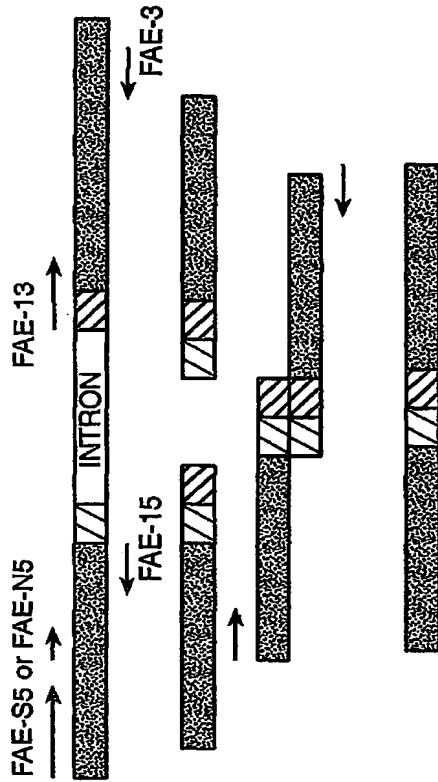


FIG._4

FAE-I3 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCATC 35-mer
 FAE-I5 GCGCCGAGGAGTGGCCGCTACCGTCAGCGCGTAGTCC 40-mer

intron position in original

Y A L T V T G H S L G A S L A A L
 GGACTACGCGCTGACCGTGACCGGCCACTCCCTCGGCGCC
 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCATC
 Y A L T V T G H A L G A S L A A L

complement, FAE-I5
 FAE-I3

FIG._5

Vector Construction

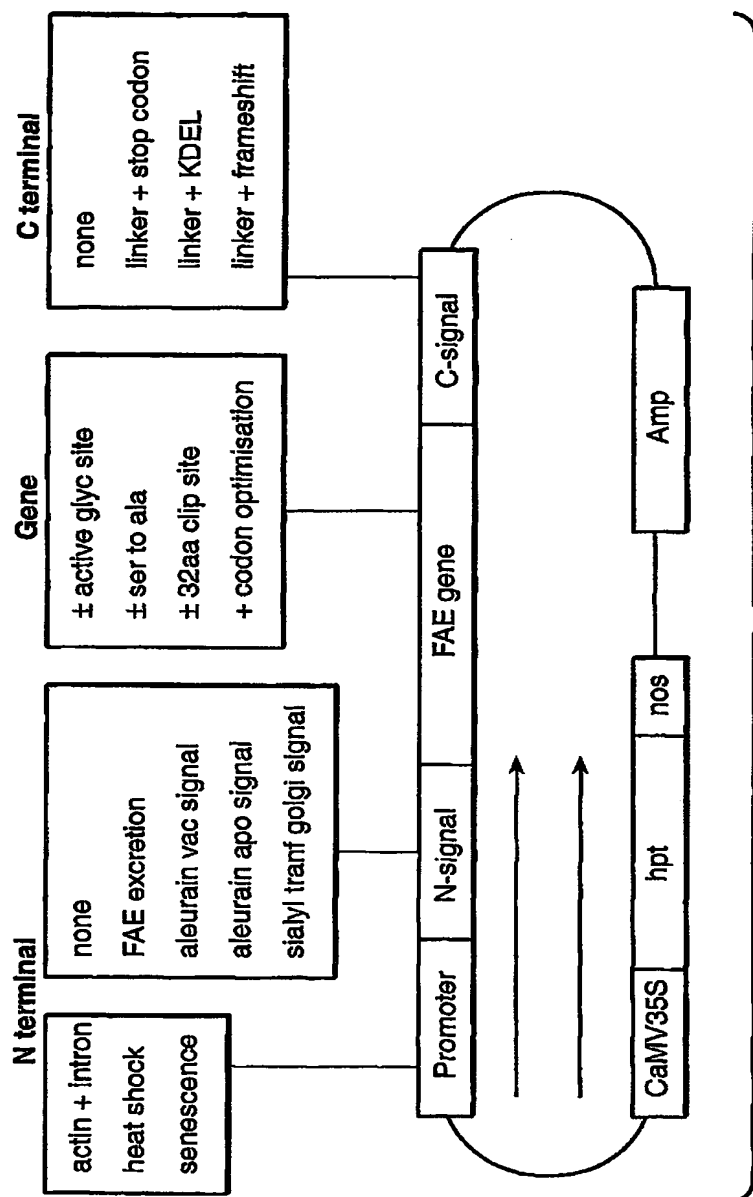
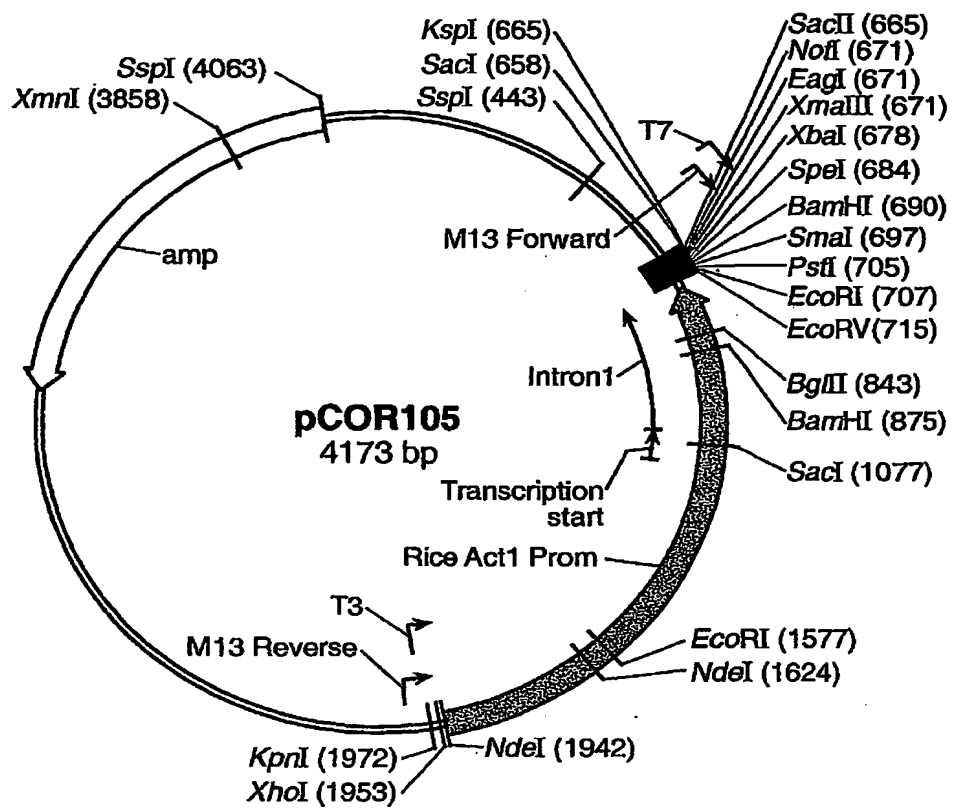
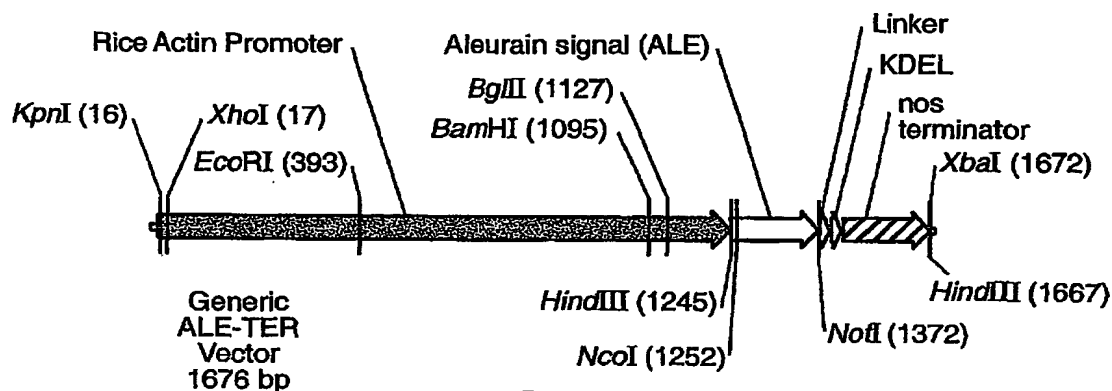


FIG._6

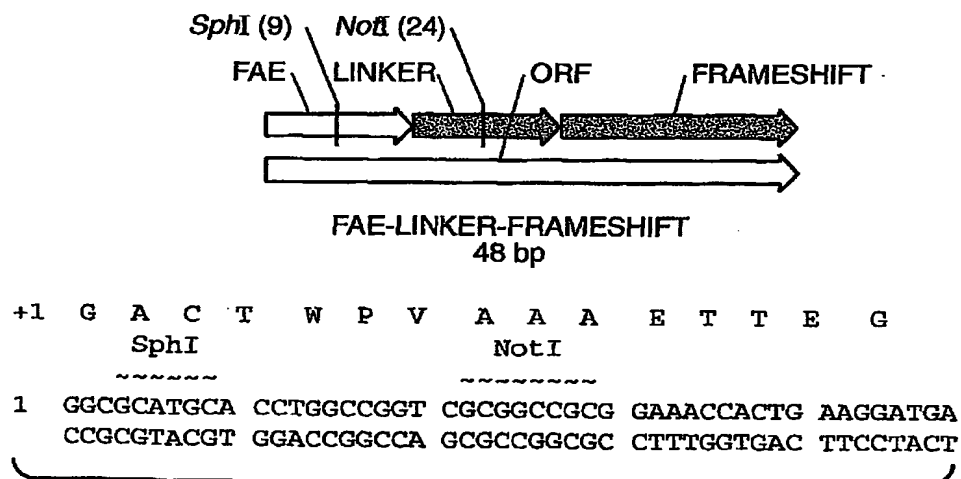
**FIG. 7**

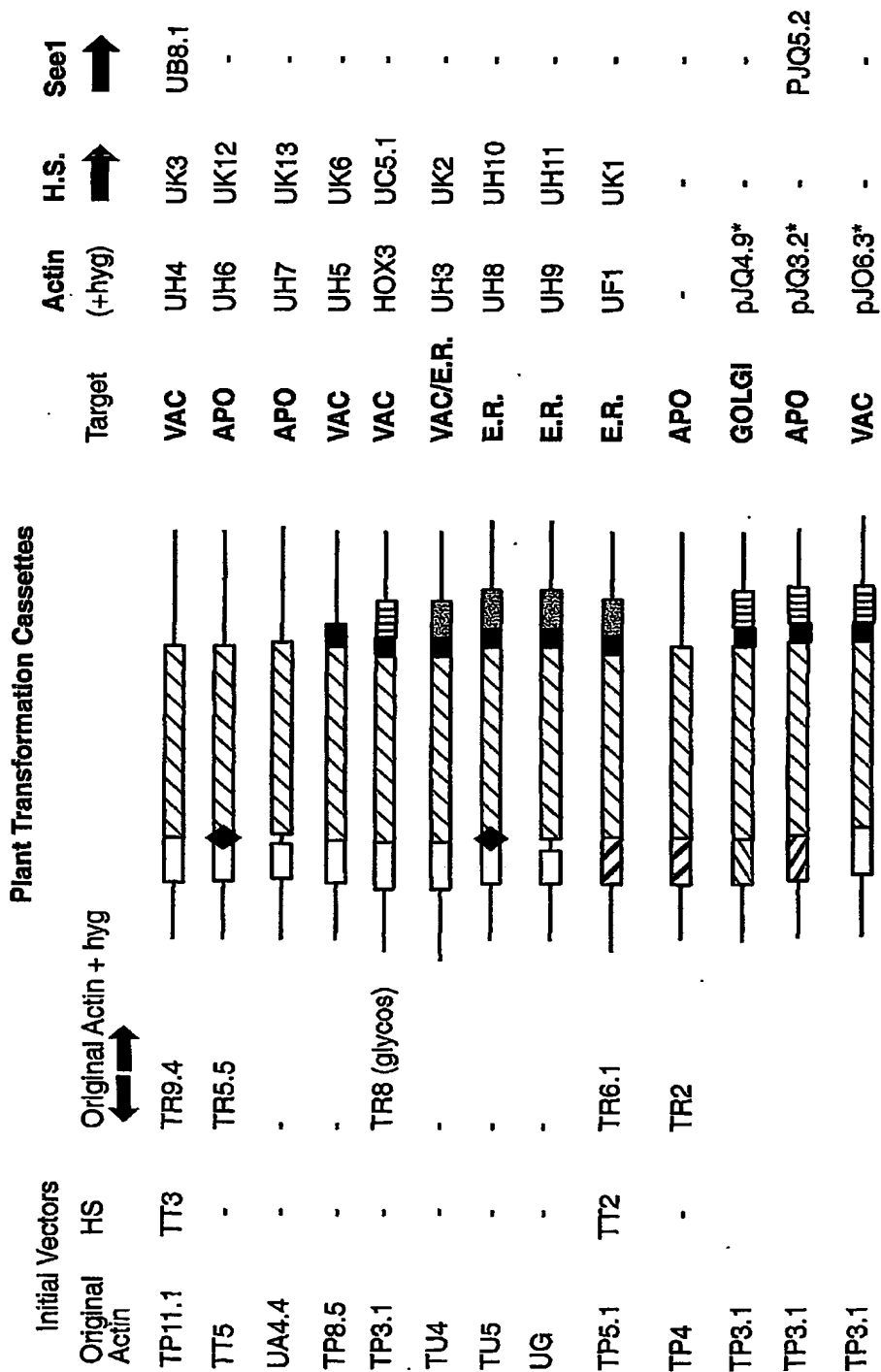
**FIG._8**

KDEL-COOH ER retention sequence

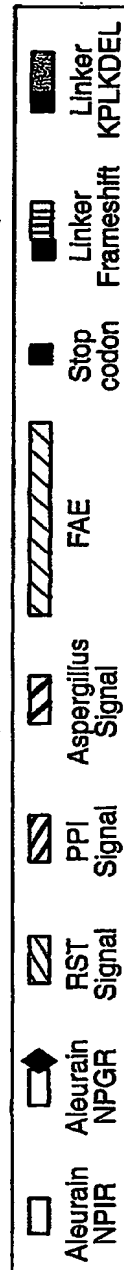
NotI

A A A K P L K D E L *
1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

FIG._9**FAE-LINKER-FRAMESHIFT Structure and Sequence****FIG._10**



* - Modified Actin Promoter (Kpn1-EcoR1 Deletion and Restored NCO Site)

**FIG. 11**

Vectors**Original Actin promoter in pCOR105**

	Target	Signal sequences	Vectors
(i)	APO	- aleurain-NPGR-FAE	pUH6, pTT5, TT5.5, pTT5.1
		- aleurain-delNPIR-FAE	pUH7, pUA4.4,
(ii)	ER	- aleurain-NPGR-FAE-linker-KDEL	pTU5, pUH8,
		- aleurain-delNPIR-FAE-linker-KDEL	pUG4, pUH9,
(iii)	VAC	- aleurain-NPIR-FAE	pTP11.1, pTR9.4, pUH4, pUK3,
(iv)	ER/VAC	- aleurain-NPIR-FAE-linker-KDEL	pTU4, pUH3,
(v)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pUA1K3, pTP3.1, pUC5.11
(vi)	VAC	- aleurain-NPIR-FAE-linker-stop	pTP8.5, pUH5
(vii)	ER	- Aspergillus signal-FAE-KDEL	pTP5.1, pTP6.1, pUF1,

Modified actin promoter (Kpn1-EcoR1 deletion and restored NCO site)

(i)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pJ06.3
(ii)	GOLGI	- RST-FAE-linker-frameshift	pJQ3.2
(iii)	APO	- PPI-FAE-linker-frameshift	pJQ4.9

Heat-shock promoter

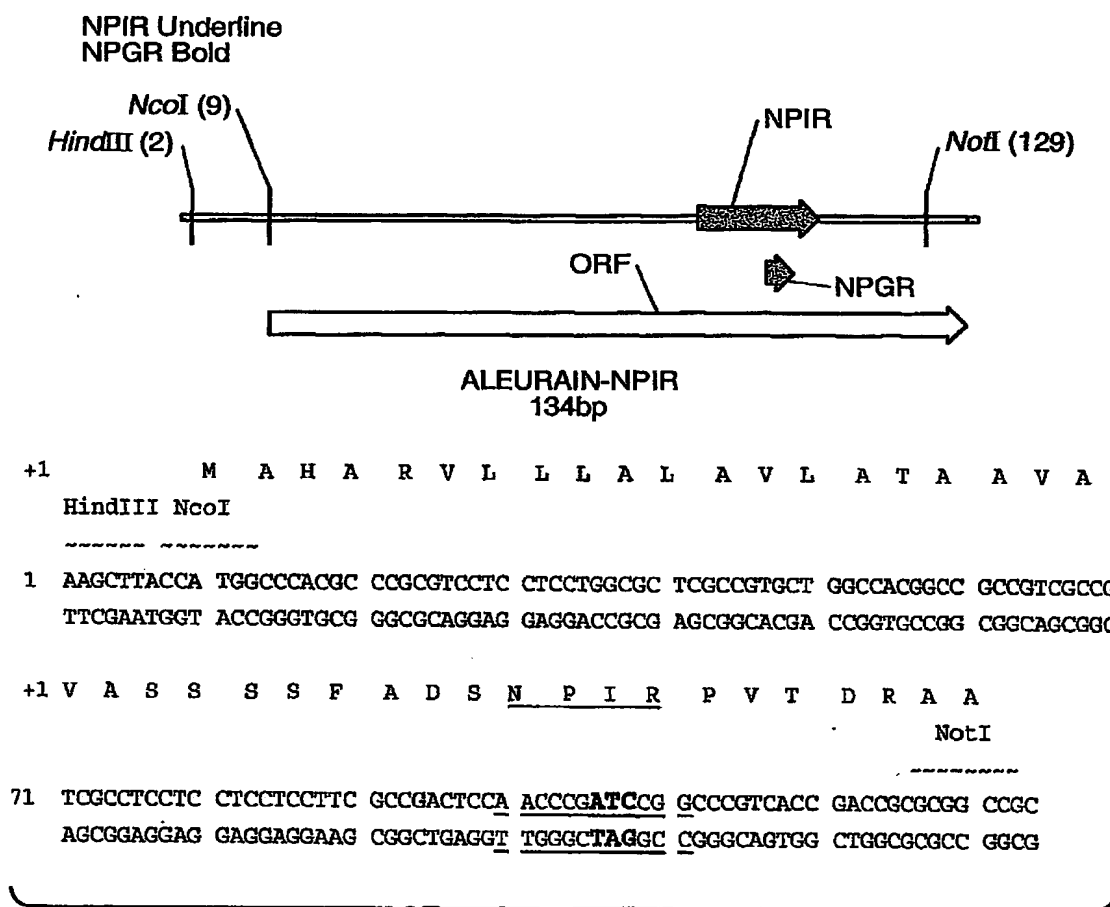
(i)	APO	- aleurain-NPGR-FAE	pUH12
		- aleurain-delNPIR-FAE	pUH13
		- Aspergillus signal-FAE	pTP4a2, pTR2.22,
(ii)	ER	- aleurain-NPGR-FAE-linker-KDEL	pUH10
		- aleurain-delNPIR-FAE-linker-KDEL	pUH11
(iii)	VAC	- aleurain-NPIR-FAE	pUK3, pTT3
(iv)	ER/VAC	- aleurain-NPIR-FAE-linker-KDEL	pUK2
(v)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pUC5.11, pHOX3
(vi)	VAC	- aleurain-NPIR-FAE-linker-stop	pUK6
(vii)	ER	- Aspergillus signal-FAE-KDEL	pUK1, pTT2

Senescence promoter

(i)	APO	- See1-PPI-FAE-linker-frameshift	pJQ5.2
(ii)	VAC	- See1-aleurain-deleted NPIR-FAE	pUB8.1

FIG. 12

ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast) Structure and Sequence



RAT SIALYL TRANSFERASE Golgi signal sequence

HindIII

 M I H T N L K X K F S L F I L V F L L F A
 1 AAGCTTACCA TGATCCACAC CAACCTCAA AAGAAGTTCT CCTCTTCAT CCTGGTCTTC CTCTCTCTCG
 . V I C V W K K G S D Y E A L T L Q A K E F Q M
 71 CCGTGATCTG CGTGTTGAAG AAGGGCTCCG ACTACGAGGC CCTCACCTTC CAAGCCAAGG AGTTCCAAAT

NotI

~~~~~

. A A

141 GCGGCGCGC

**FIG.\_14**

## POTATO PROTEASE INHIBITOR II Apoplast signal sequence

HindIII  
 -----  
 M X V H K E V N F V A Y L L I V L G L L L  
 1 AAGCTTACMA TGGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTACCTCCT GATCGTSCCTC  
 GGCCTCCTCT

NcoI

~~~~~

. L V S A M E H V D A K A C T X E C G N L
 G F G .

71 TGCTCGTSTC CGCCATGGAG CAGTGGAGG CCAAGGCCCTG CACCCCKGAG TCGGCGAACC
 TCGGCTTCGG

NotI

~~~~~

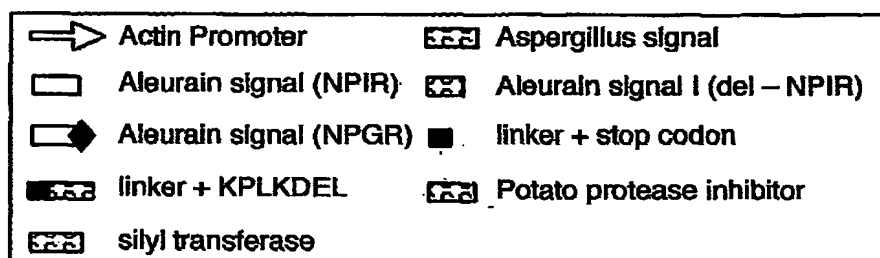
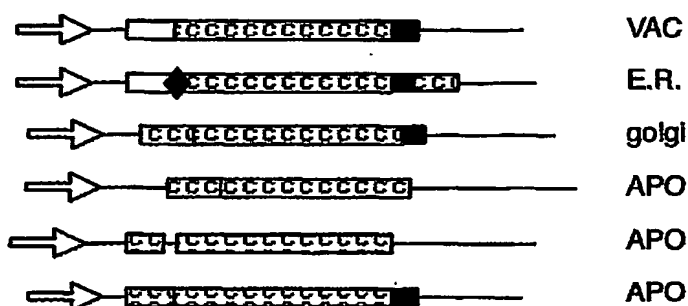
. I C P A A A

141 CATCTGCCCG GCGGCGCGC

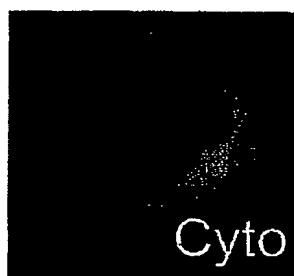
**FIG.\_15**

# Targeting Expression of gfp to Different Cell Compartments

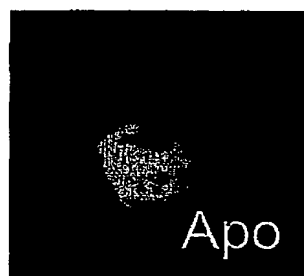
## Actin Promoter Targeting Vectors



**FIG. 16A**



**FIG. 16B**



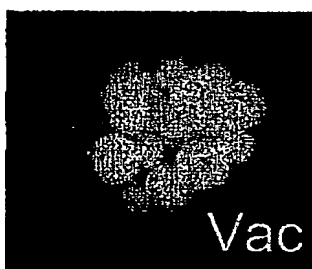
**FIG. 16C**



**FIG. 16D**



**FIG. 16E**

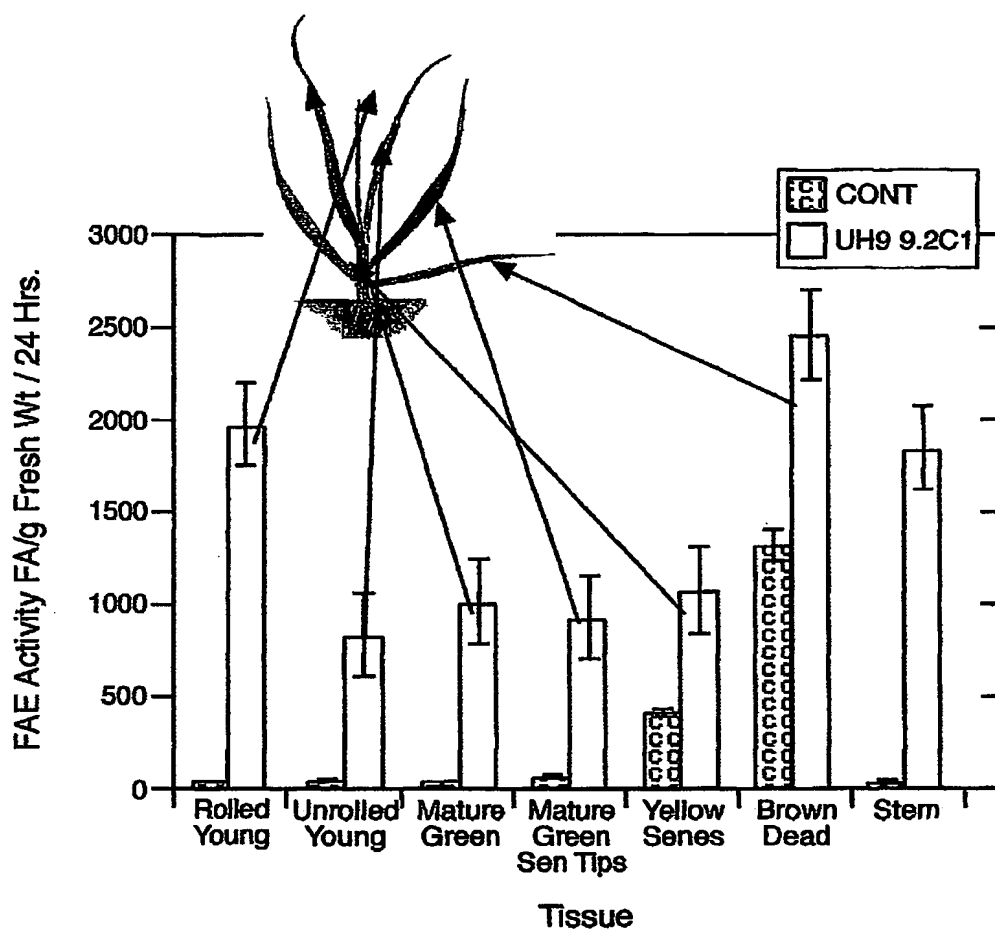


**FIG. 16F**



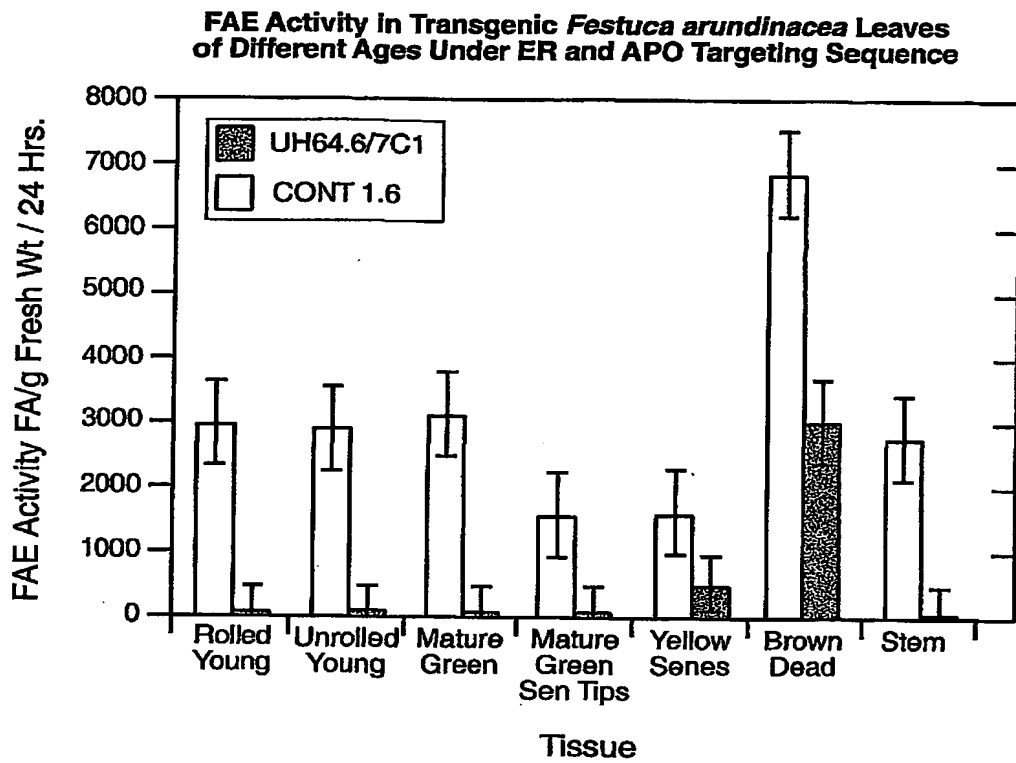
**FIG. 16G**

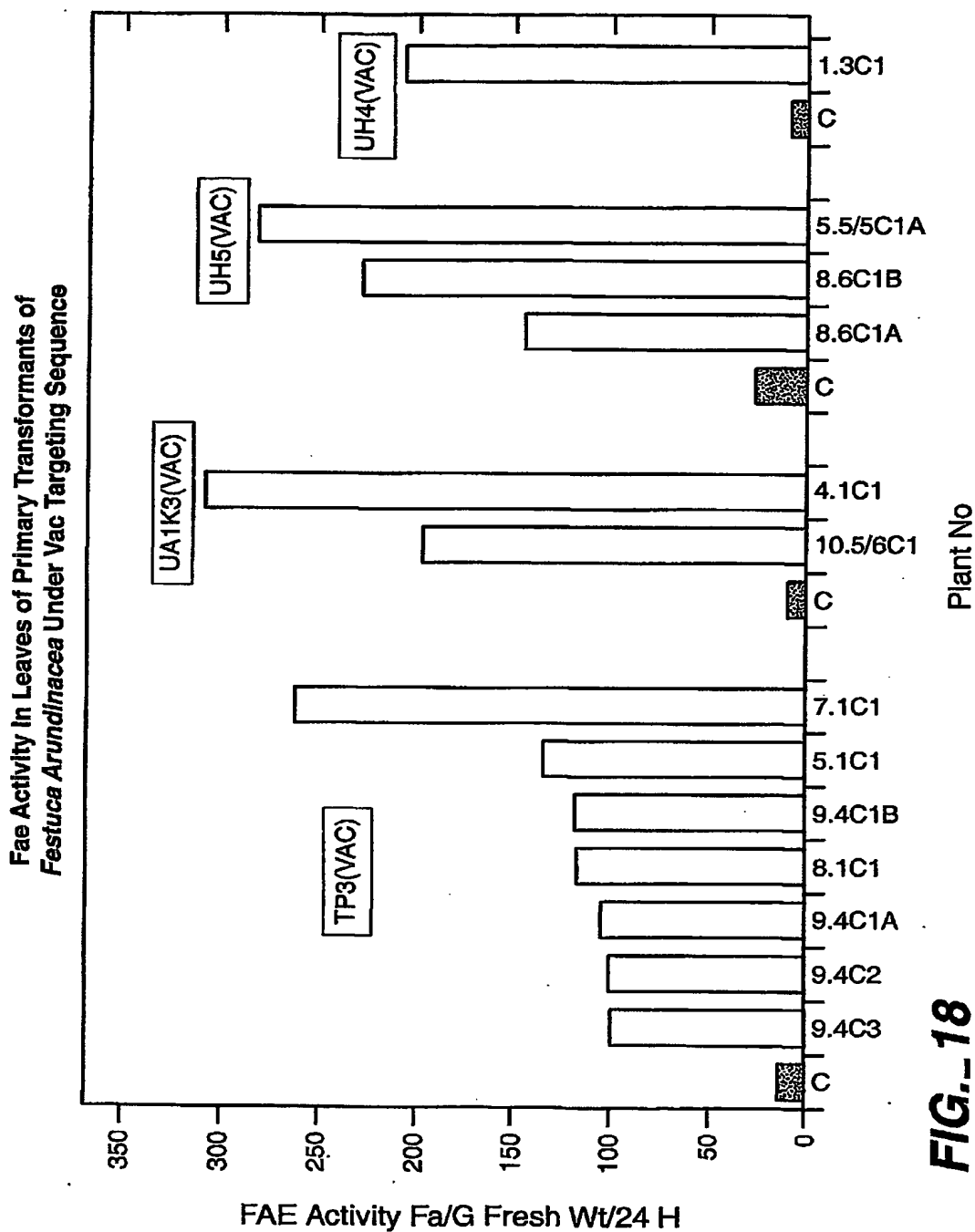
**F AE Activity in Transgenic *Festuca arundinacea* Leaves  
of Different Ages Under ER and APO Targeting Sequence**

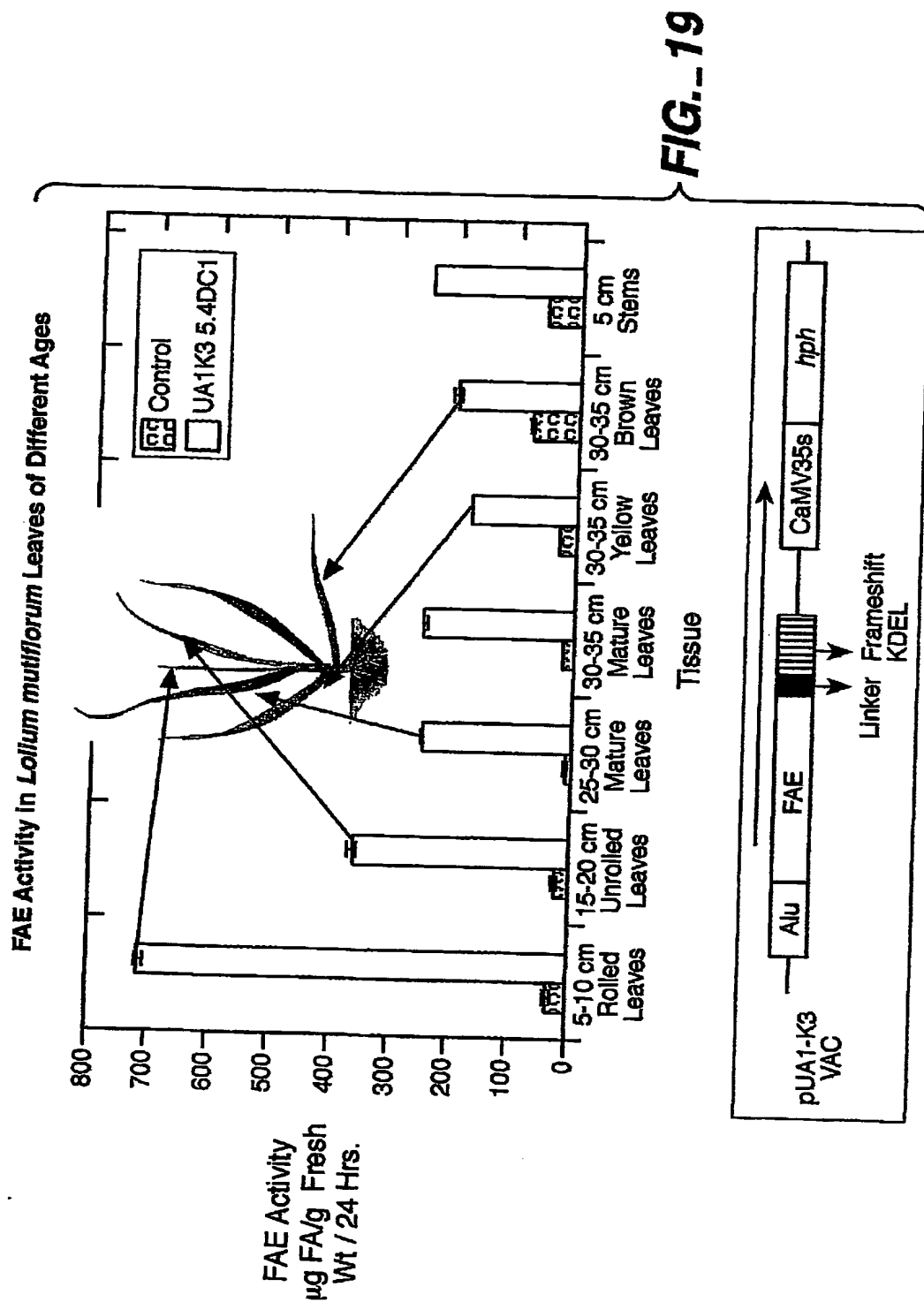


**FIG. 17A**



**FIG. 17B**





FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

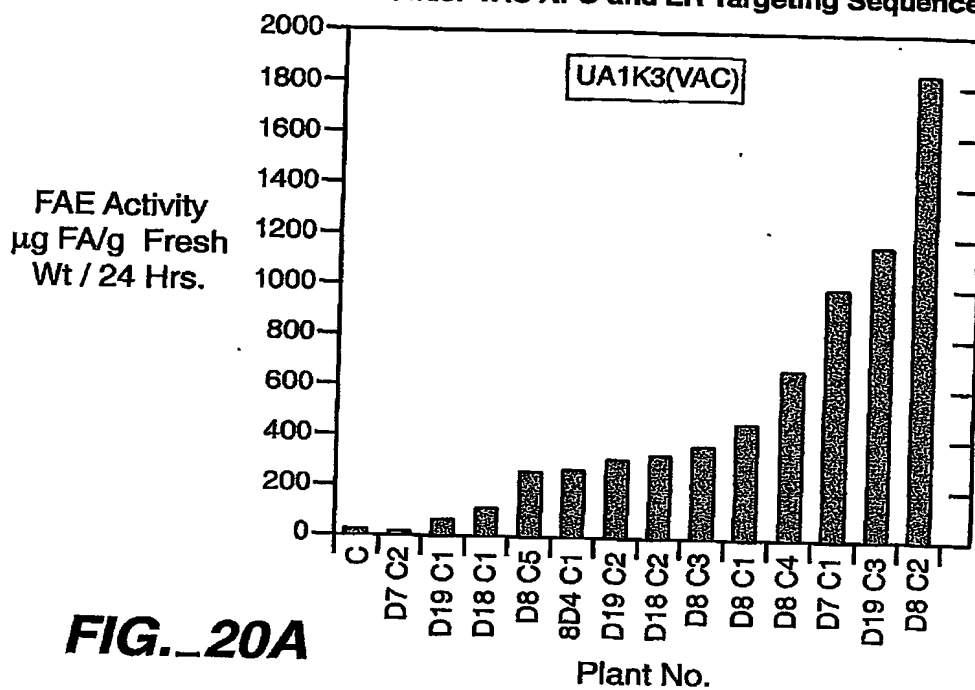


FIG. 20A

FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

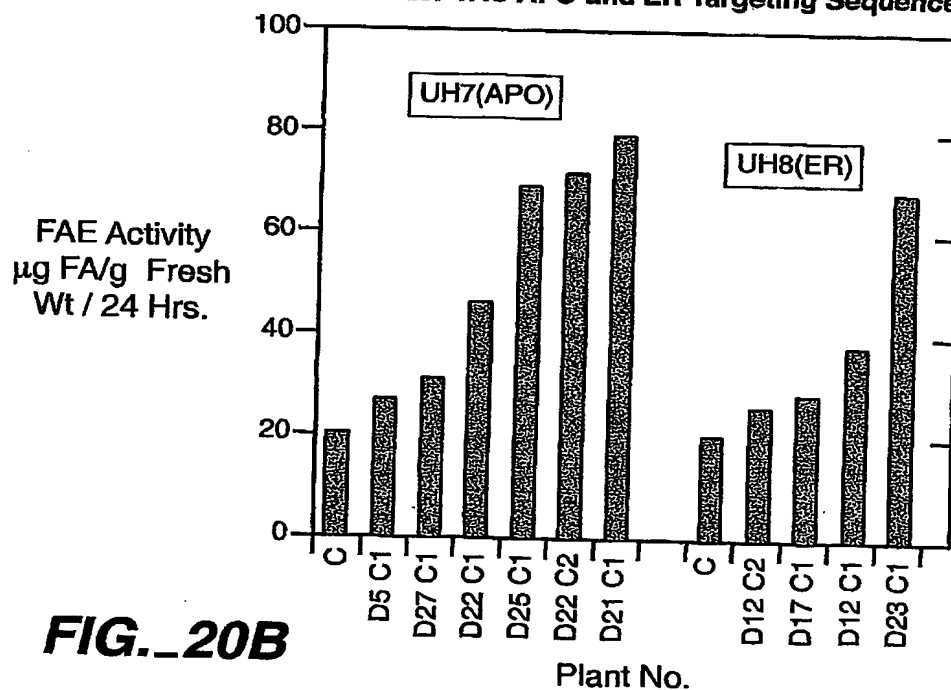


FIG. 20B

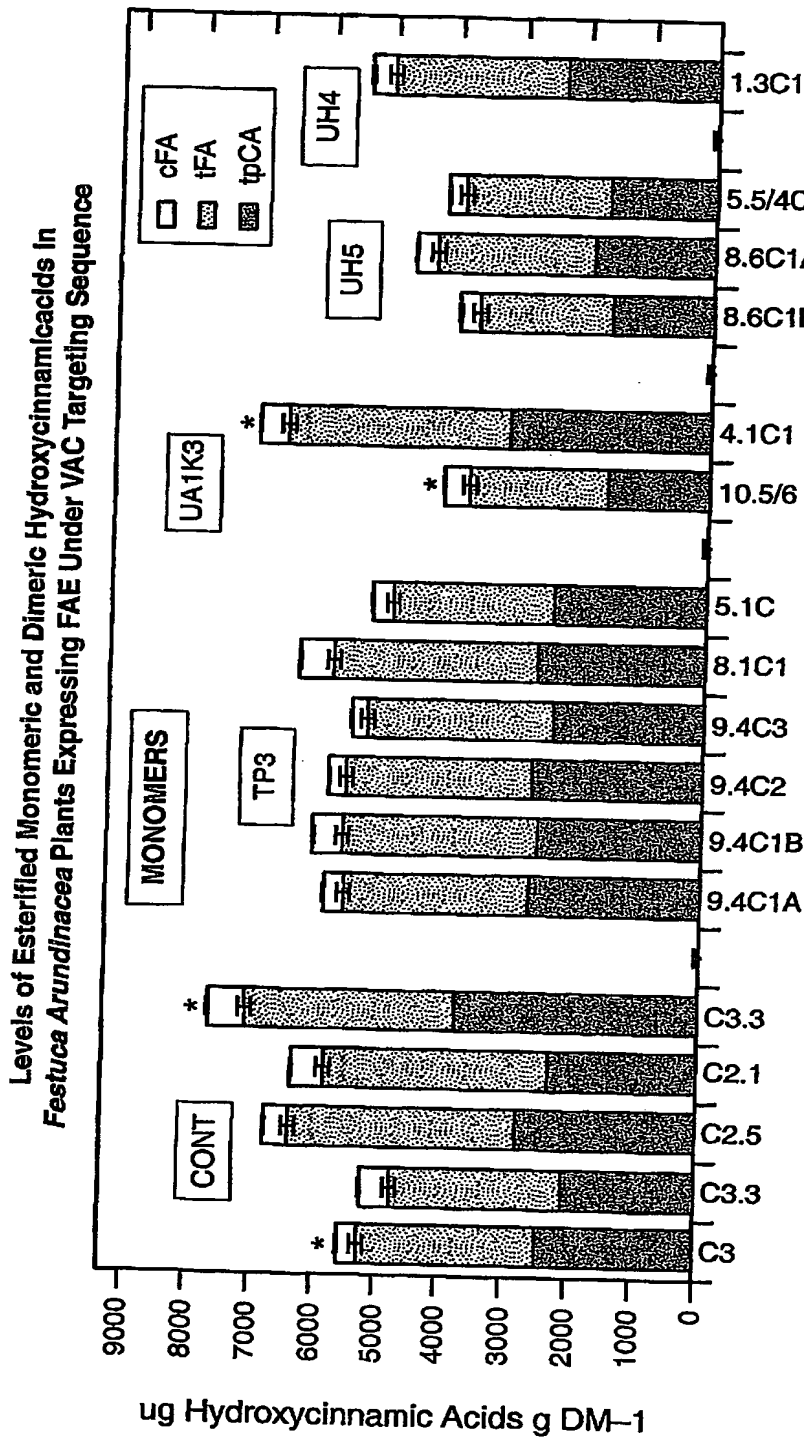


FIG. 21A

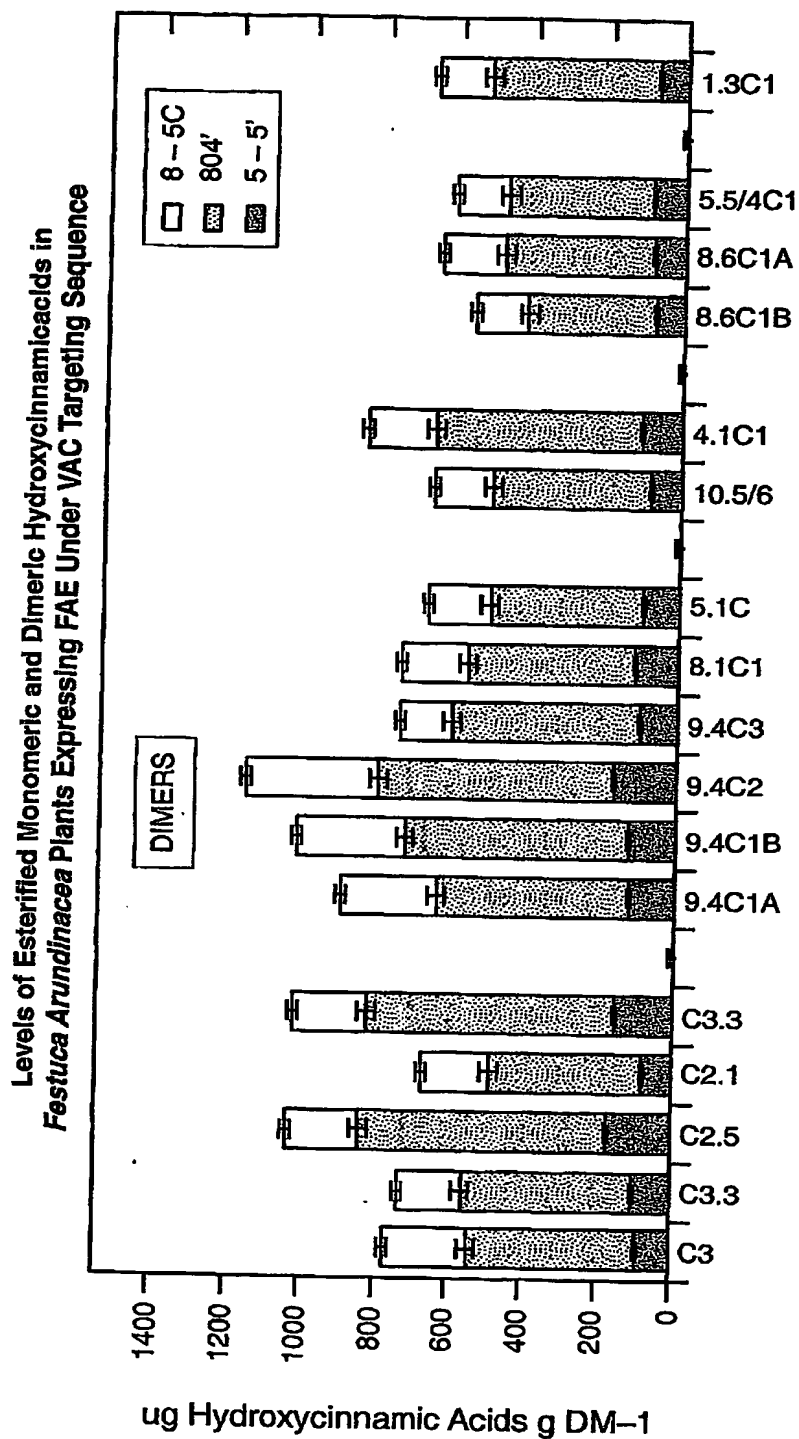


FIG. 21B

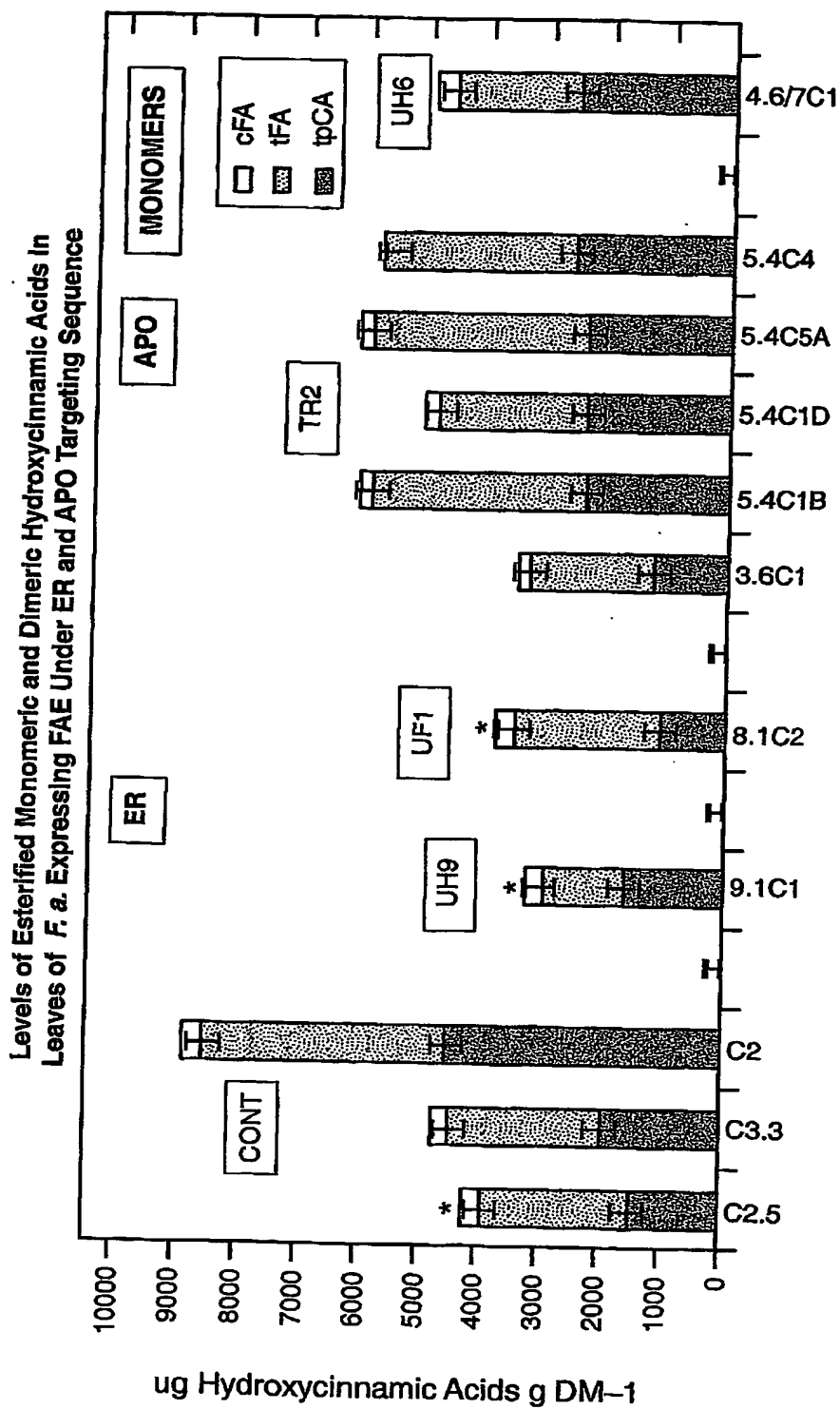
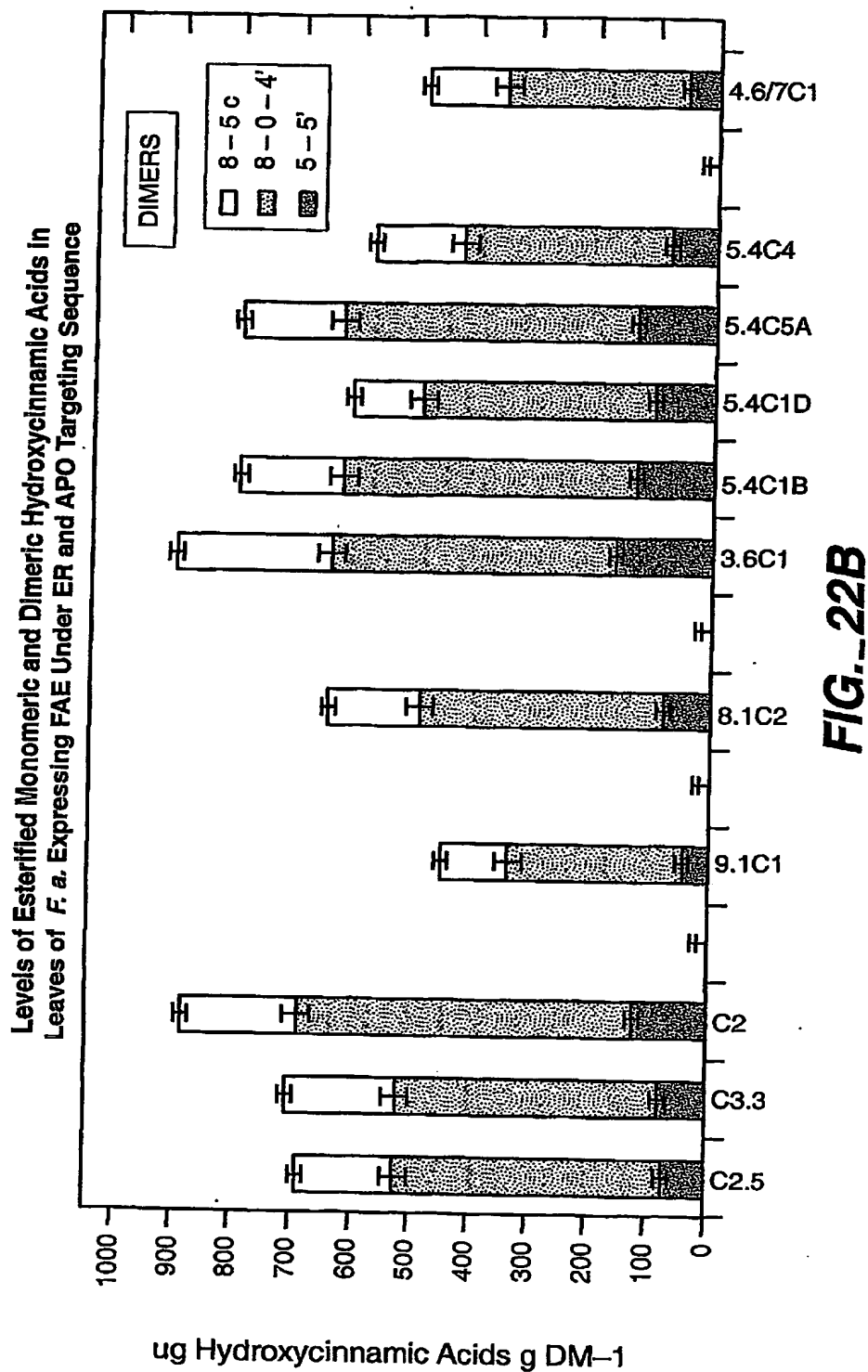
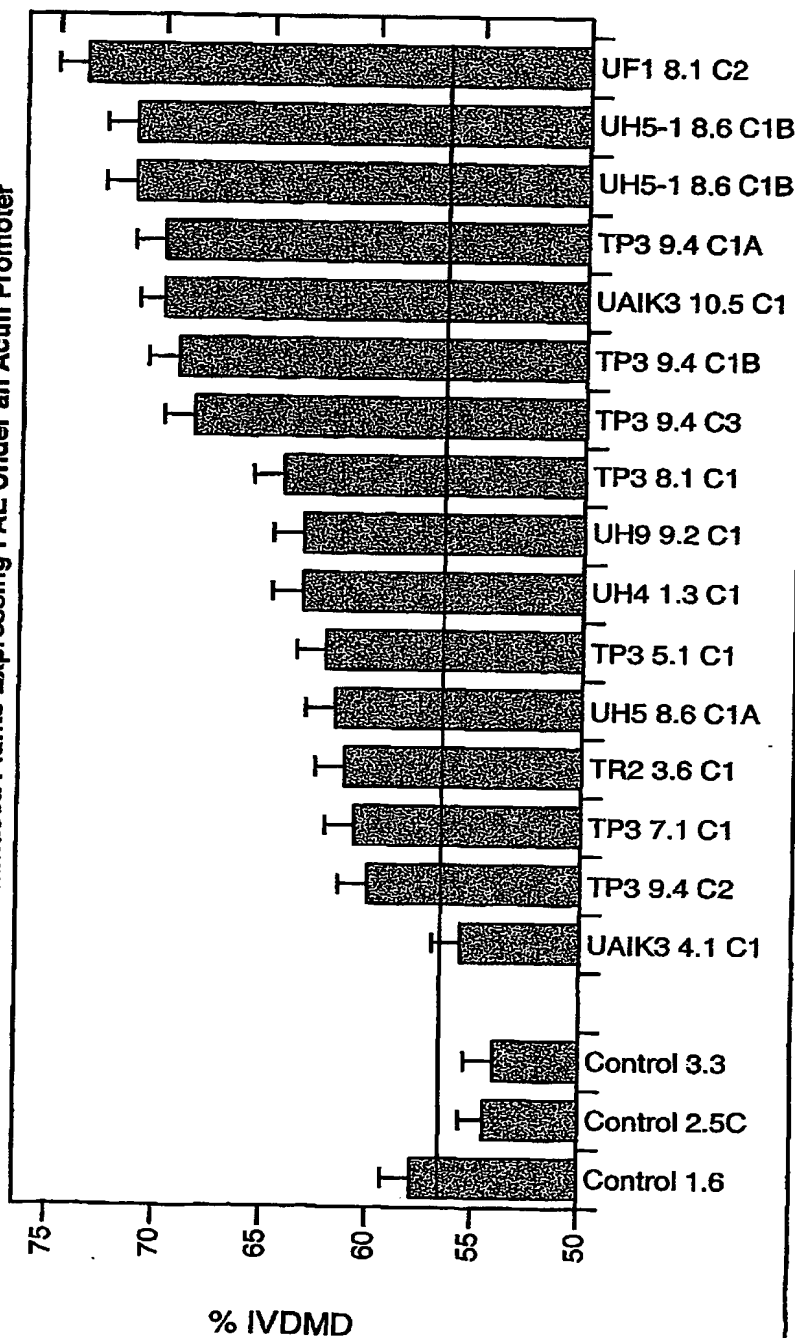


FIG. 22A





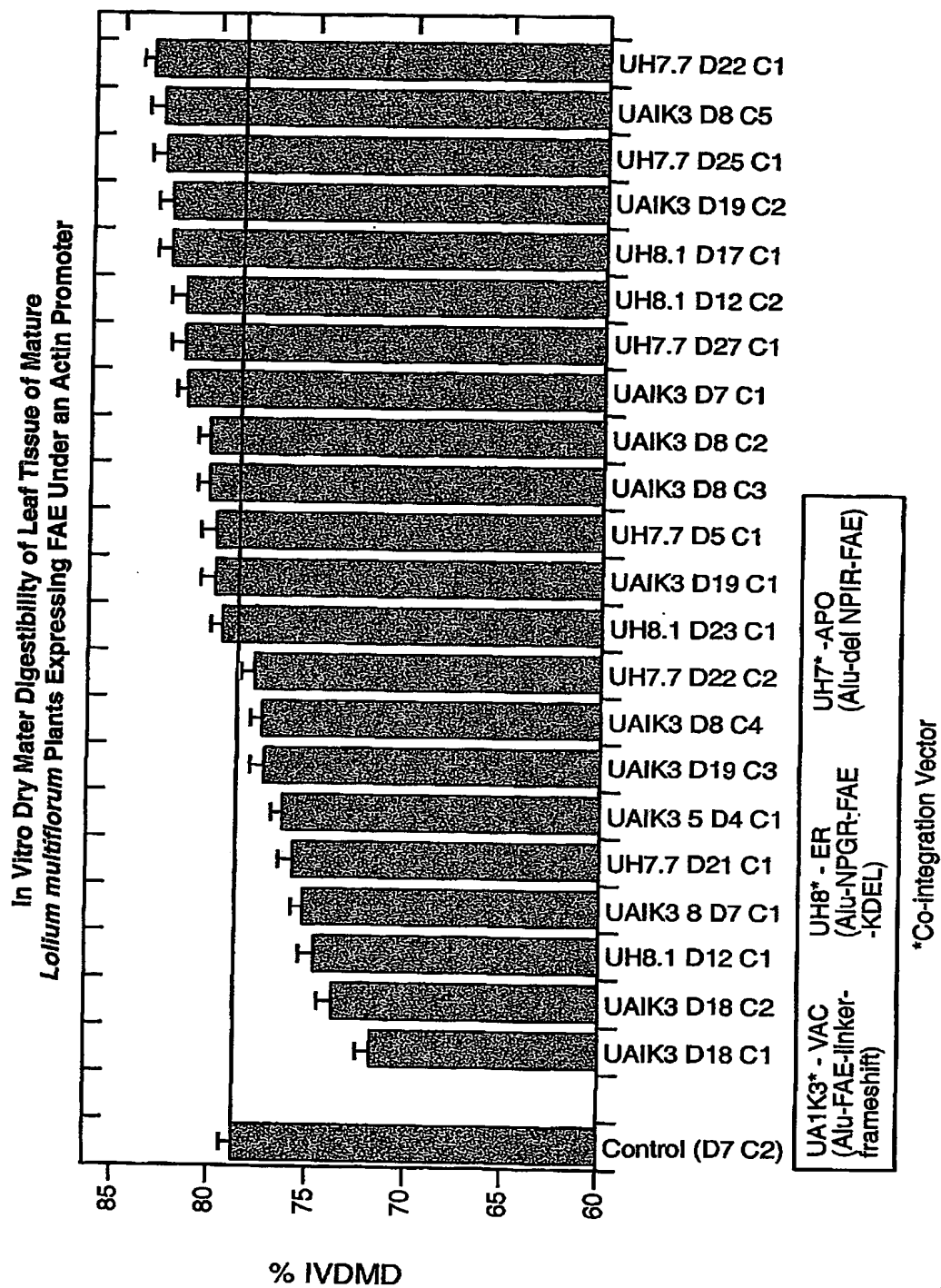
In Vitro Dry Matter Digestibility of Leaf Tissue of Mature  
*Festuca arundinacea* Plants Expressing FAE Under an Actin Promoter



|                                                 |                                              |                        |                                         |                                       |                                             |                                        |
|-------------------------------------------------|----------------------------------------------|------------------------|-----------------------------------------|---------------------------------------|---------------------------------------------|----------------------------------------|
| UA1K3* - VAC<br>(Alu-FAE-linker-<br>frameshift) | TP3 - VAC<br>(Alu-FAE-linker-<br>frameshift) | TR2 - APO<br>(Asp-FAE) | UH5* - VAC<br>(Alu-FAE-<br>linker-stop) | UH4* - VAC<br>(Alu-FAE +<br>CaMV-Hyg) | UH9* - ER<br>(Alu-del NP1R-<br>linker-KDEL) | UF1* - ER<br>(Asp-FAE-<br>linker-KDEL) |
|-------------------------------------------------|----------------------------------------------|------------------------|-----------------------------------------|---------------------------------------|---------------------------------------------|----------------------------------------|

\*Co-integration Vector

**FIG. 23**

**FIG. 24**

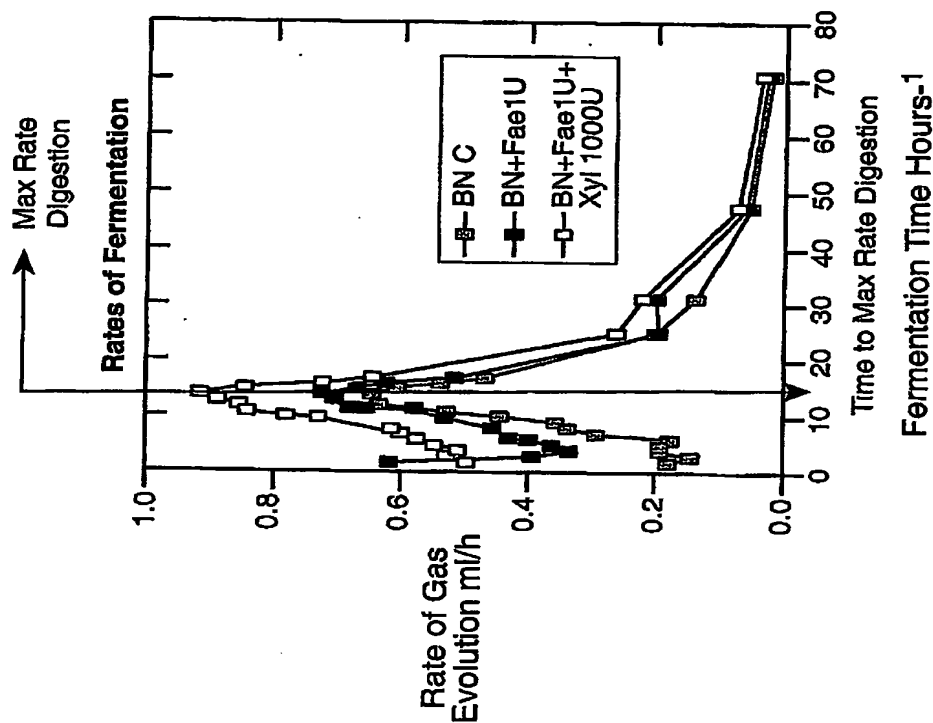


FIG. 25B

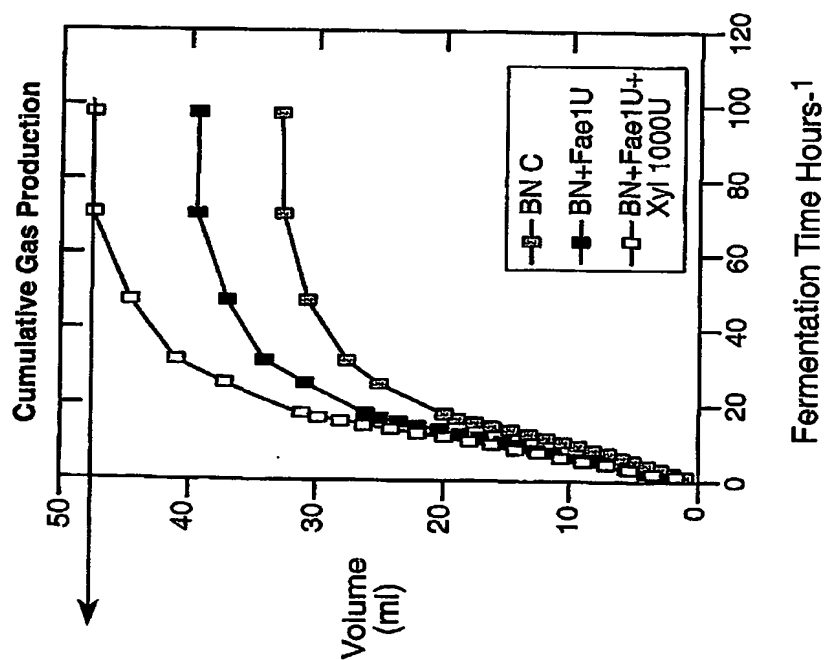
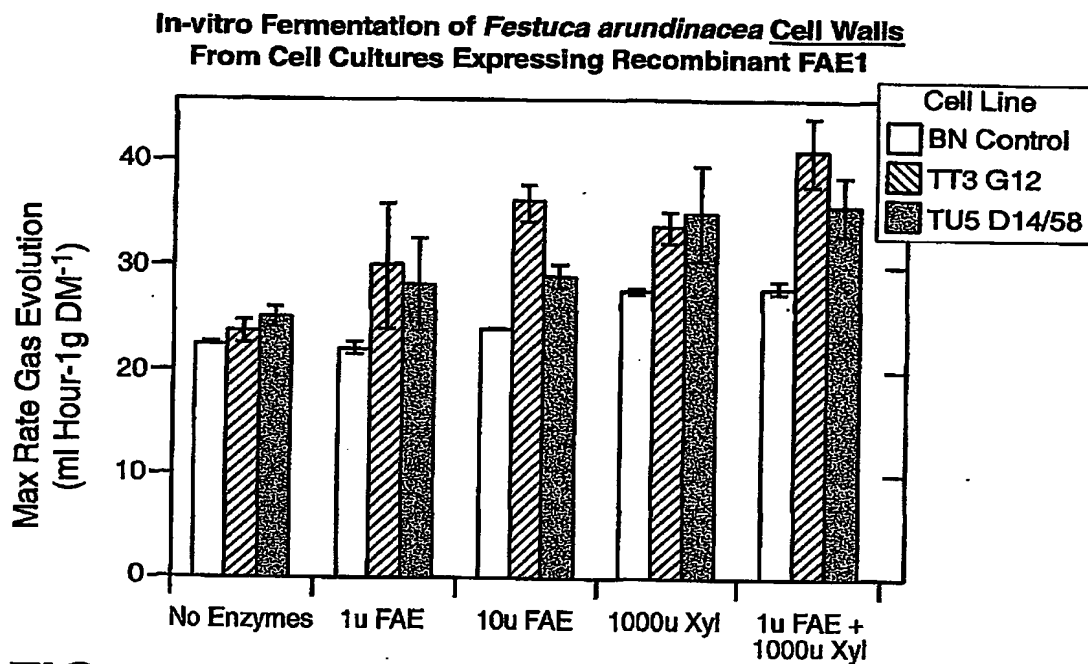
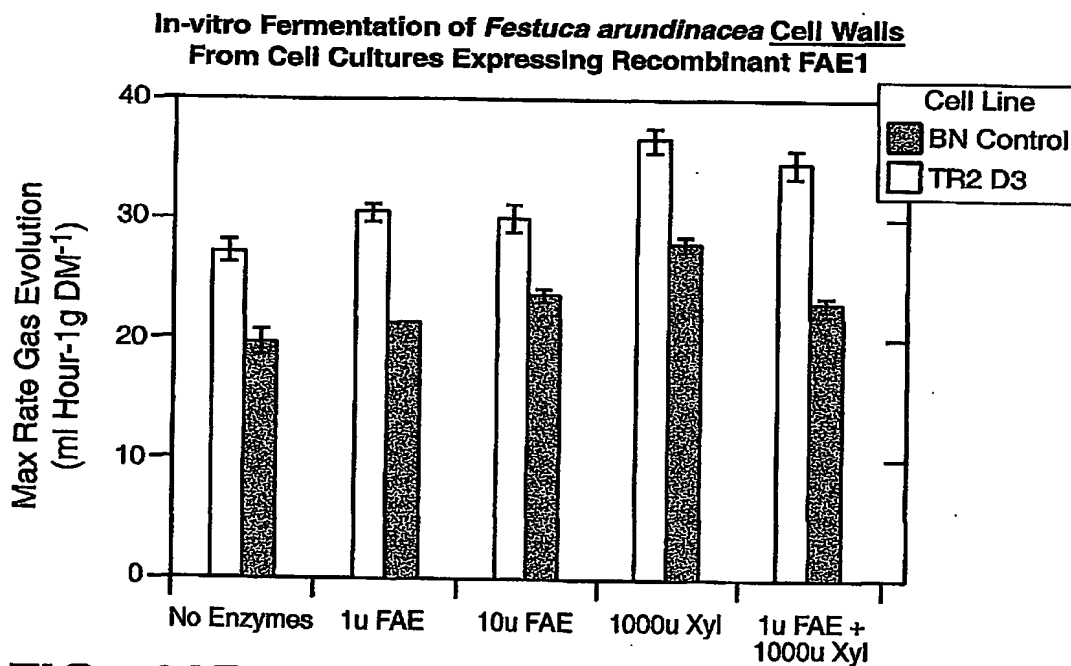


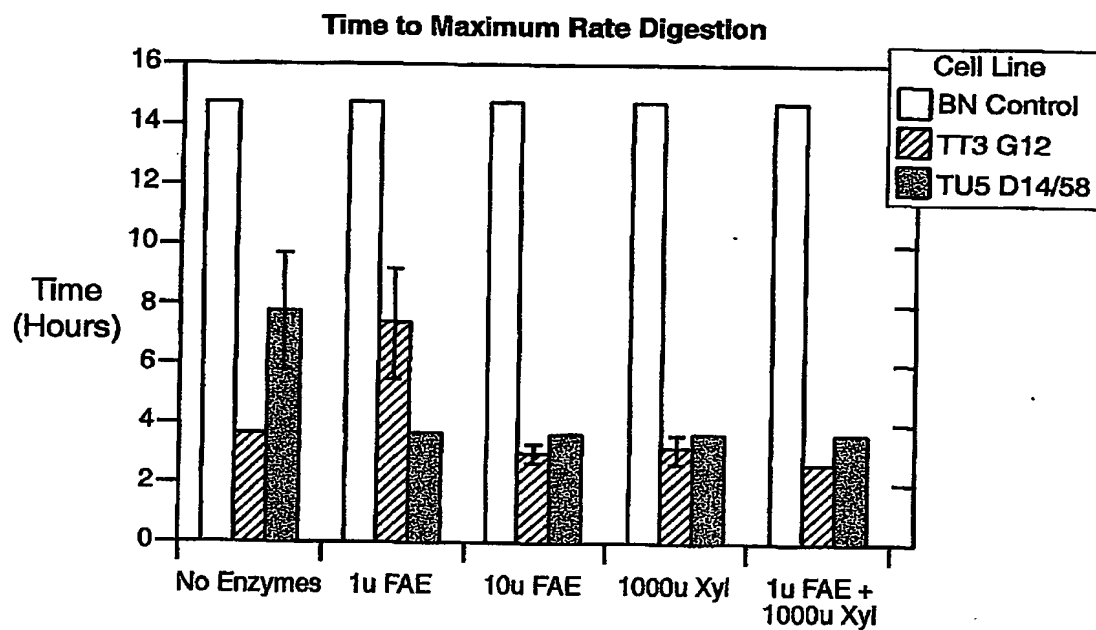
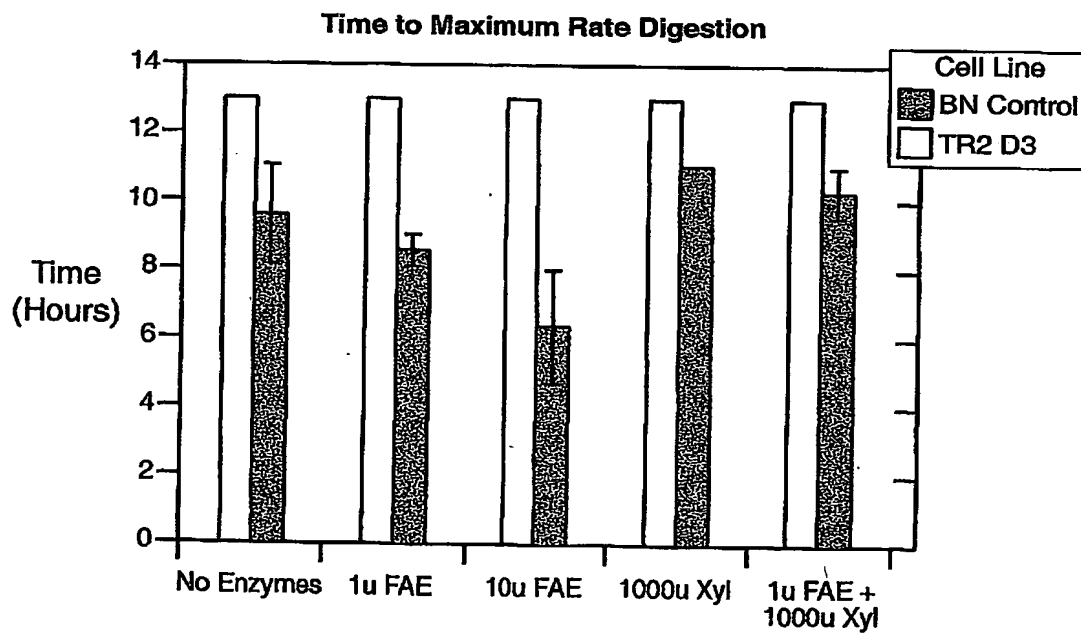
FIG. 25A

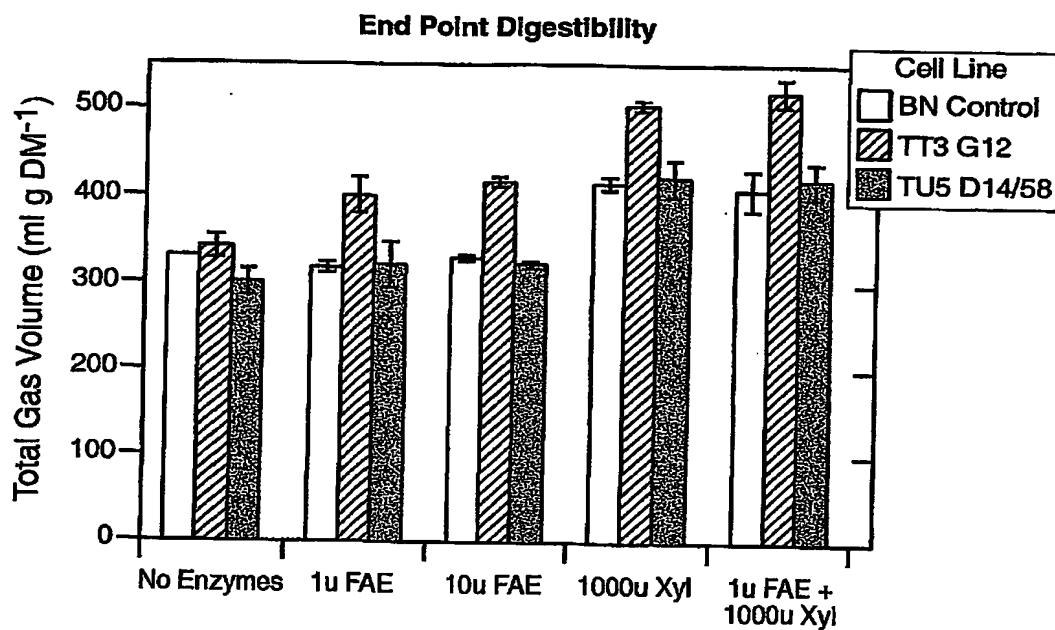
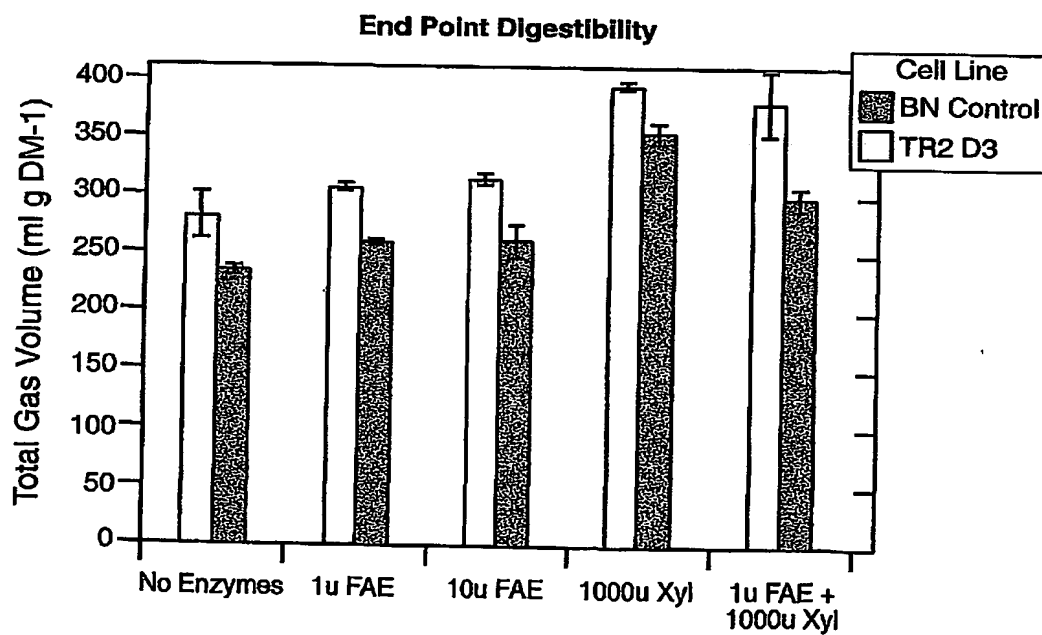


**FIG. 26A** Maximum Rate of Digestion

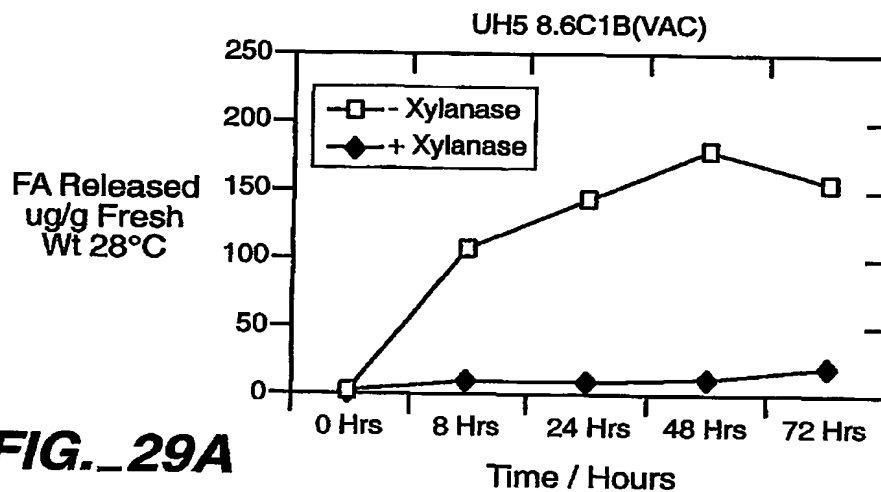


**FIG. 26B** Maximum Rate of Digestion

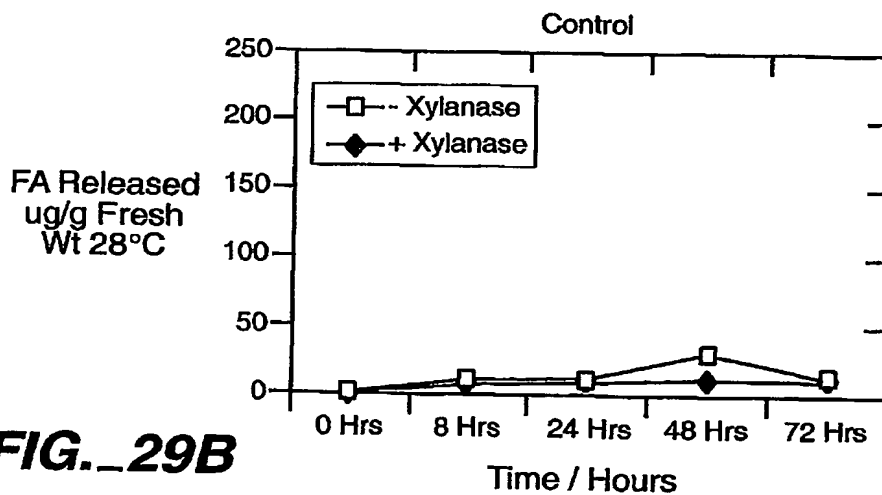
**FIG. 27A****FIG. 27B**

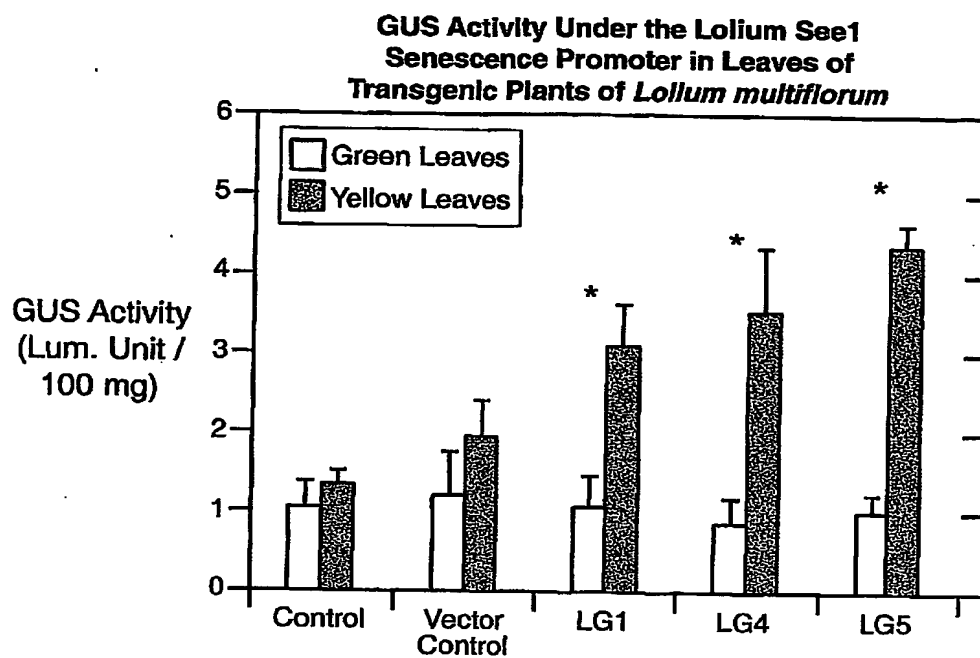
**FIG.\_28A****FIG.\_28B**

**Kinetics of FAE Activity by Ferulic Acid Release  
from Cell Wall under Self Digestion in *Festuca arundinacea*  
and Stimulation by Xylanase**

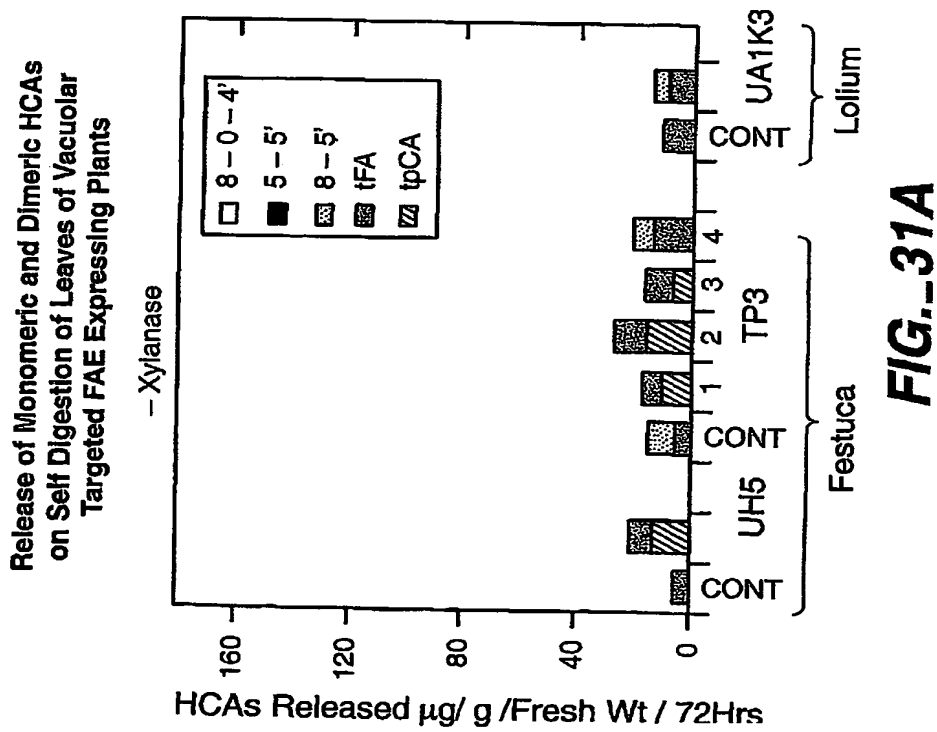
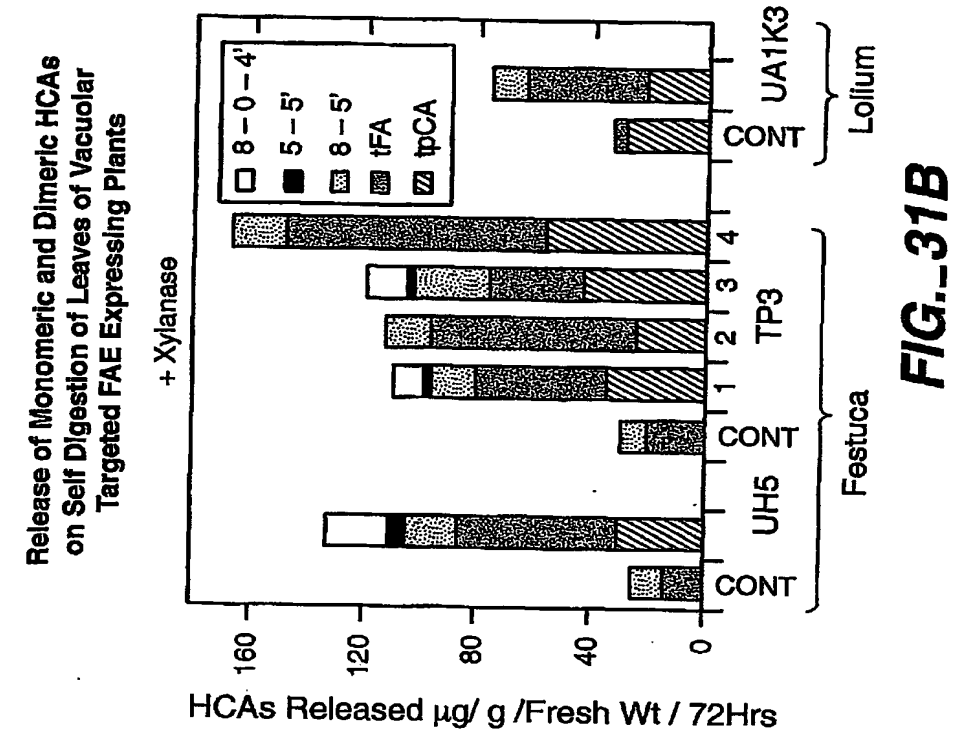
**FIG.\_29A**

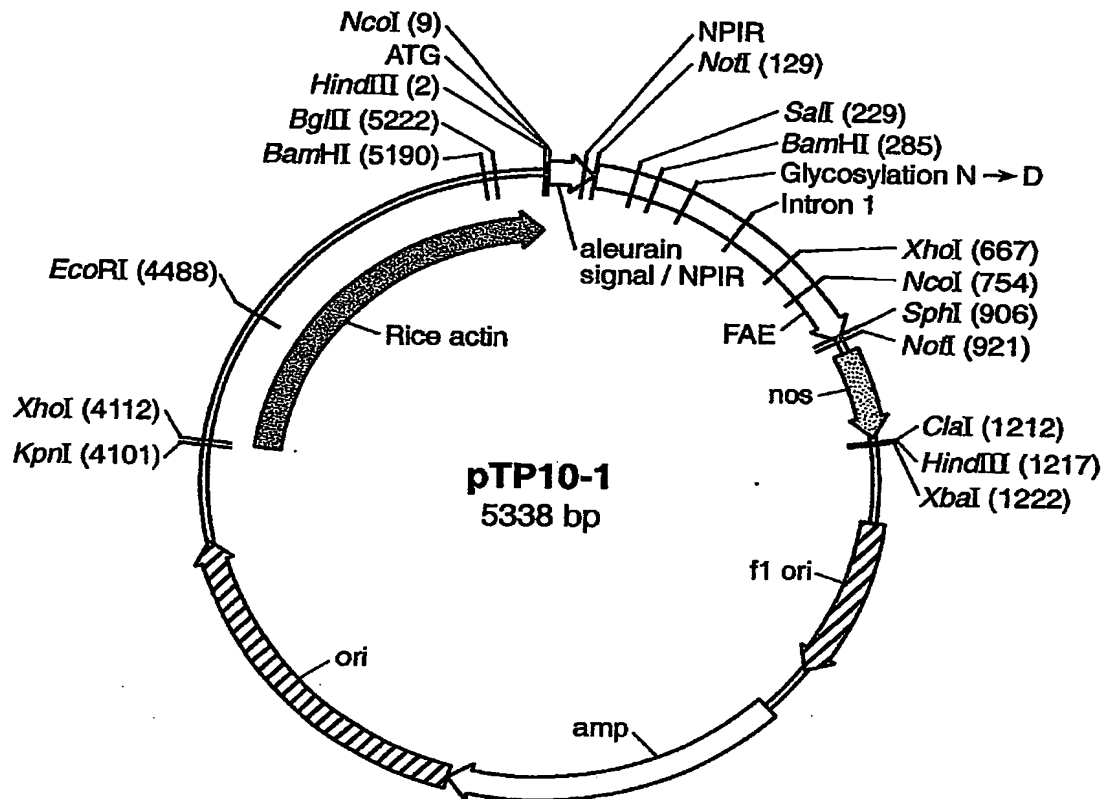
**Kinetics of FAE Activity by Ferulic Acid Release  
from Cell Wall under Self Digestion in *Festuca arundinacea*  
and Stimulation by Xylanase.**

**FIG.\_29B**

**FIG.\_30**





**FIG.\_32A**

NcoI  
 HindIII  
 M A H A R V L L L A L A V L A T A A V A V  
 1 AAGCTTACCA TGGCCACGC CCGGTCTC CTCTGGCGC TCGCGTGCT GGCACGGCC GCGTCGCGG  
  
 NPIR  
 NotI  
 . A S S S S F A D S N P I R P V T D R A A A S T .  
 71 TCGCCTCCTC CTCCTCCTTC GCCGACTCA ACCCGATCCG GCCGTCCACC GACCGCGCG CCGCCTCCAC  
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A  
 141 GCAGGCAATC TCCGAAGACC TCTACAGCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC  
  
 SalI  
 D L C N I P S T I I K G E K I Y N S Q T D I N G  
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGAG AAATTACAA TTCTCAACT GACATTACG  
  
 BamHI  
 . W I L R D D S S K E I I T V F R G T G S D T N .  
 281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTCCGT GGCACGTGTA GTGATACGAA  
  
 Glycosylation  
 . L Q L D T D Y T L T P F D T L P Q C N G C E V  
 351 TCTACAACTC GATACTGACT ACACCTCAG CCCTTTCGAC ACCCTACAC AATGCAACGG TTGTGAAGTA  
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
 421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGATC GCTTGTCAA CAGCAGGTTA  
 . Q Y P D Y A L T V T G H X L G A S L A A L T A .  
 491 GCCAGTATCC GGAATACGGG CTGACCGTGA CCGGCCACKC CCTCGCGCC TCCCTGGCGG CACTCAGTGC  
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q  
 561 CGCCCAAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG

FIG.\_32B

XhoI  
 ~~~~~  
 631 GCCTTGGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA

NcoI
 ~~~~~  
 701 CTGATGCCAA CGACGGCATC CCAAGCTGC CCCGGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACGCGG ATGAGTGA GTGCTGTGAG

SphI  
 ~~~~~  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGGCG GCATGCACCT

NotI
 ~~~~~  
 911 GCGCGGTGCG GCGCGCGGAA ACCACTGAAG GATGAGTGT AAAGAAGCAG ATCGTTCAA CATTGGCAA  
 981 TAAAGTTTCT TAAGATTGAA TCCTGTGGCC GGTCTTGGCA TGATTATCAT ATAAATTCG TTGAATTACG  
 1051 TTAAGCAATG AATAATTAAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGGA TTAGAGTCCC  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAATT ATCGCGCGCG

HindIII  
 ~~~~~  
 1191 GGTGCACTTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCGC CCTATAGTGA
 1261 GTCGTATTAC GCGCGTAC TGGCCGTCTG TTTACAACGT CGTACTGGG AAACCCCTGG CGTTACCCAA
 1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGTGGC GTAAATAGGA AGAGGCCCGC ACCGATCGCC
 1401 CTTCCCAACA GTTGGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGCGCGG

Clal
 ~~~~~  
 XbaI  
 ~~~~~

FIG. 32C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTTCTTC
 1541 CTTTCTTTTC TCGCCACGTT CCGCGGCTTT CCCCCTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC
 1611 GATTAGTGC TTACAGGCAC TTACAGGCAC AAAAAGCTGA TTAGGGTGTG GTTACACGTA GTGGGCCATC
 1681 CCGCTGATAG ACGGTTTTC CACTCAACC TATCTCGGC ATTTTAAAGG GATTTGCCG ATTTGTCGCT
 1751 ACTGGAACAA AATGAGCTG AATTAACAA AATTTAAGC GATTTTAAAC AAAATATTA CATTCAAATA
 1821 ATTGTTTAA CTTTTCGCG CATGAGACAA TAACCTGTAT AAATGCTTCA ATATATTTGA AAAAGGAAGA GTATGAGTAT
 1891 TTAGGTGGCA CTTTTCGCG CAGTACGCT CAGTACGCT TTTTTCGCTT CAGTACGCTT TTTTTCGCTT TCACCCAGAA
 1961 TGTATCCGCT TCAACATTTT CAGTACGCT CAGTACGCT TTTTTCGCTT CAGTACGCTT TTTTTCGCTT TCACCCAGAA
 2031 TCAACATTTT CAGTACGCT CAGTACGCT TTTTTCGCTT CAGTACGCTT TTTTTCGCTT CAGTACGCTT TCACCCAGAA
 2101 ACGCTGGTGA AAGTAAAGA GATCTTGA GATCTTGA CAGTGGGTG CAGGAGTGG CAGGAGTGG CAGGAGTGG
 2171 ACAGCGGTAA GATCTTGA GATCTTGA CAGTGGGTG CAGGAGTGG CAGGAGTGG CAGGAGTGG CAGGAGTGG
 2241 GCTATGTGG CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2311 CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2381 TATGAGTGC TGCATTAAC ACCGCTTTT TGCATTAAC CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2451 GAAGGAGCTA ACCGCTTTT TGCATTAAC CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2521 CTGAATGAAG CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2591 AACTATTAAC TGCATTAAC CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2661 AGTTCAGGA CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2731 GAGCGTGGT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2801 ACACGACGG GAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT CAGTACGCT
 2871 TAAGCATGG TAACTGTGAG ACCAAGTTT GATCTTTT GATCTTTT GATCTTTT GATCTTTT GATCTTTT
 2941 TTTAAAGGA TCTAGGTGAA GATCTTTT GATCTTTT GATCTTTT GATCTTTT GATCTTTT GATCTTTT
 3011 TCCACTGAG CTCAGACCC GATCTTTT GATCTTTT GATCTTTT GATCTTTT GATCTTTT GATCTTTT
 3081 CTGCTGCTTG CAAACAAA AACCACCGT CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3151 CTTTTCCTGA AGGTAACTGG CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3221 TAGGCCACCA CTTCAAGAAC TCTGTAGC CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3291 TGCTGCCAGT GCGGATAAGT CTTGTAGC CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3361 CGGTCGGCT GAACGGGGG TTTGTGCTA CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3431 ACCTACAGCG TGAGCTATGA GAAAGCGCA CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3501 CCGCAGGGTC GGAACAGGAG AGCGCAGCAG AGCGCAGCAG AGCGCAGCAG AGCGCAGCAG AGCGCAGCAG
 3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CTTACGCT CTTACGCT CTTACGCT CTTACGCT CTTACGCT
 3641 AAAACGCCAG CAACGCGGCT TTTTACCGT TTTTACCGT TTTTACCGT TTTTACCGT TTTTACCGT
 3711 TGGGTTATCC CTTGATTTCTG TGGATTAACG TTTTACCGT TTTTACCGT TTTTACCGT TTTTACCGT
 3781 CGAACGACCG AGCGCAGCGA GTCAGTGGC GAGGAGCGG AAGAGCGCC AAGAGCGCC AAGAGCGCC
 3851 CCGCGCGTTG GCCGATTCAT TAAATGCAGT GGCACGACAG GTTTCGCGC GTTTCGCGC GTTTCGCGC

FIG.-32D

```

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC

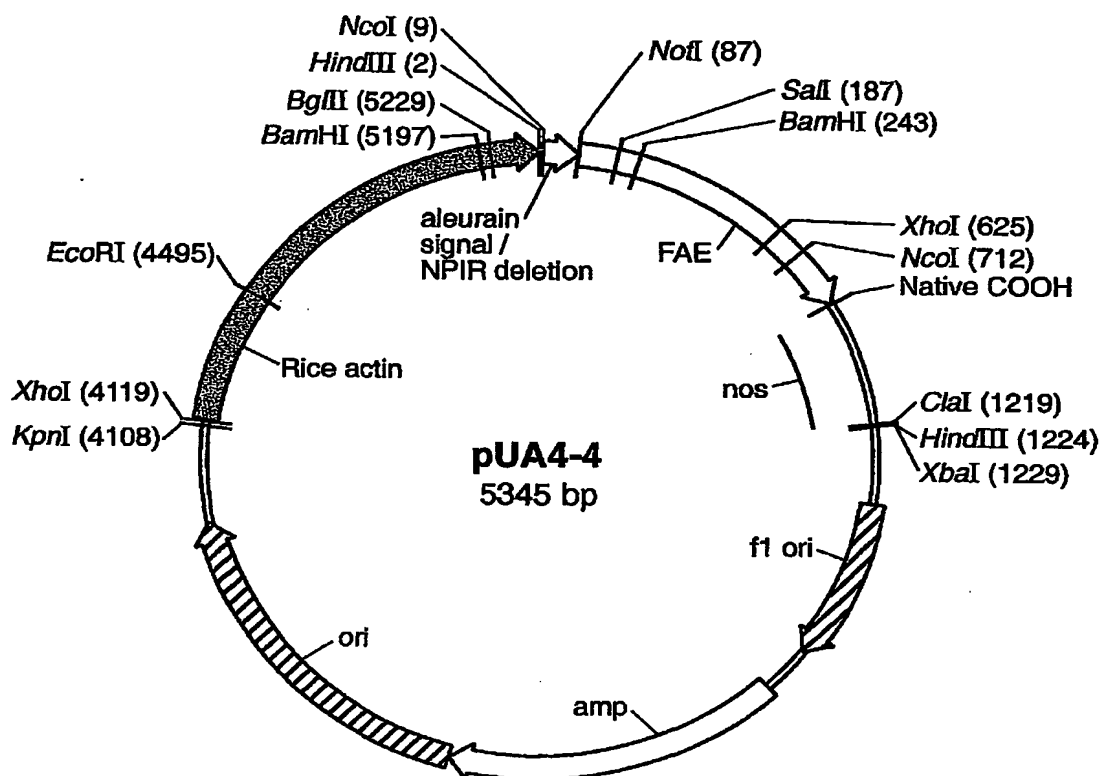
                               KpnI
                               ~~~~~
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCCC CTCGAGGTCA TTCATATGCT
4131 TGAGAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGGTA AGATTACCTG GTCAAAAGTG AAAACATCAG
4201 TTAAAAGGTT GTATAAGTAA AATATCGGTA ATAAAGGTTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAT TTGAGGATGT TTGTCGGTAC TTTGATACGT CATTTTGTGA TGAATTGGTT TTTAAGTTTA
4341 TTCGCGATTT GGAATGCAAT ATCTGTATTT GAGTCGGTTT TTAAGTTTCG TGCCTTTGTA AATACAGAGG
4411 GATTTGTATA AGAATATATCT TTAATAAAC CATATGCTAA TTTGACATAA TTTTGTGAAA AAATATATAT

                               EcoRI
                               ~~~~~
4481 TCAGGCGAAT TCCACAAATGA ACAATAATAA GATTAAATA GCITGGCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAGATA AACTTAGACT CAAAACATTT ACAMAAACAA CCCCTAAAGT CCTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC AACCCACCCC AGTGCAGCCA
4691 ACTGGCAAAAT AGTCTCCACC CCCGGCACTA TCACCGTGAG TTGTCCGCAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAAA AAGAAAGAAA AAAAAAGAAA AGAAAAACAG CAGGTGGGTC CGGTCGTGG GGGCCGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGGC CCGGCCCTCC CTCCGCTTCC AAAGAAACGC CCCCCTCGC
4901 CACTATATAC ATACCCCCC CTCTCTCTCC ATCCCCCAA CCTACCAACC ACCACCAACC CCACCTCTC
4971 CCCCCTCGCT GCCGACGAC GAGCTCCTCC CCCCTCCCC TCAGCCGCGG CCGGTAACCA CCCCCTCCCT
5041 CTCCTCTTTC TTCTCTCGTT TTTTCTTTCG TCCTCGGTCTC GATCTTTGGC CTCTGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCGC GGGAGGGGCG GGAATCTCGG GCTGGCGTCT CCGGGCGTGA

                               BamHI
                               ~~~~~
5181 GTCGGCCCCG ATCCTCGCGG GGAATGGGC TCTCGGATGT AGATCTTCTT TCTTTCTTCT TTTTGTGTA
5251 GAATTTGAA CTCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTTCATGAT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTGTAGC

```

FIG. 32E

**FIG._33A**

```

NGOI
~~~~~
HindIII
~~~~~
      M A E A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACAGC CCGGTCCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCCGTGCGCG

      ~~~~~
      . A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCCTCCTC CCGCGCGCC GCCTCCAGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT

      ~~~~~
      . A T I S Q A A A Y A D L C N I P S T I I K G E K
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTGCA CTATTATCAA GGGAGAGAAA

      ~~~~~
      I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTTACAATT CTCAACTGA CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATAATCACCG

      . F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAACTAC ACCCTCAGC CTTTCGACAC

      . L P Q C N G C E V H G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCACGGTT GTGAAGTACA CCGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA

      V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAACA GCAGGTAGC CAGTATCCGG ACTACGGCT GACCGTGACC GGCCACKCCC

      . G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGGCGCCTC CCTGGCGCA CTCACTGCCG CCCAGCTGTC TCGACATATC GACAACATCC GCCTGTATAC

      ~~~~~
      . F G E P R S G N Q A F A S Y M N D A F Q A S S
561 CTTGCGCGAA CCGCGCAGCG GCAATCAGCG CTTGCGGTCG TACATGAACG ATGCTTTCCA AGCCTCGAGC

      P D T T Q Y F R V T H A N D G I P N L P F V E Q
631 CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAAGC ACGGCATCCC AAACCTGCCC CCGGTGGAGC

      ~~~~~
      . G Y A H G G V E Y W S V D P Y S A Q N T F V C .

```

FIG._33B

701	AGGGGTACGC	CCATGGCGGT	GTAGAGTACT	GGAGCGTTGA	TCCTTACAGC	GCCCAGAAC	CATTGTCTG
	T G D	E V Q C	E A	Q G G	Q G V N	N A H	T T Y
771	CACGTGGGAT	GAAGTGCAGT	GCTGTGAGC	CCAGGGCGGA	CAGGGTGTGA	ATAATGCGCA	CACGACTTAT
	F G M T	S G A	C T W	*			
841	TTTGGGATGA	CGAGCGGAGC	CTGTACATGG	TGATCAGTCA	TTTTCAGCCTC	CCCGAGTGT	CCAGGAAAGA
911	TGGATGTCTT	GGAGAGGGGG	CCGCTTAACC	ACTGAAGGAT	GAGCTGTAAA	GAAGCAGATC	GTTCAAAACAT
981	TTGGCAATAA	AGTTTCTTAA	GATTGAATCC	TGTTGCCGGT	CTTGCCGATGA	TTATCATATA	ATTTCTGTTG
1051	AATTACGTTA	AGCATGTAA	AATTAAACATG	TAATGCTAAG	CGTTATTTAT	GAGATGGGTT	TTTATGATTA
1121	GAGTCCCGCA	ATTATACATT	TAATACGCGA	TAGAAACAA	AATATAGCGC	GCAAACTAGG	ATAAATTATC
				HindIII			
			Clal	XbaI			
1191	GCGCGCGGTG	TCATCTATGT	TACTAGATCG	ATAAGCTTCT	AGAGCGGCCG	GTGGAGCTCC	AATTCGCCCT
1261	ATAGTGAGTC	GTATTACGG	CGCTCACTGG	CCGTCTGTTT	ACAACGTCTG	GACTGGGAAA	ACCTTGGCGT
1331	TACCCAACTT	AATCGCTTG	CAGCACATCC	CCCTTTCGCC	AGCTGGCGTA	ATAGCGAAGA	GGCCCGCACC
1401	GATCGCCCTT	CCCAACAGTT	GCGCAGCCTG	AATGGCGAAT	GGGACGGGCC	CTGTAGCGGC	GCATTAAAGC
1471	CGCGGGGTGT	GGTGGTTACG	GCGCAGGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTCCG
1541	TTTCTTCCCT	TCCTTCTCG	CCACGTTGCG	CGGCTTCCC	CGTCAAGCTC	TAAATCGGG	GCTCCCTTTA
1611	GGGTTCCGAT	TTAGTGCTTT	ACGGCACCTC	GACCCCAAA	AACTTGATTA	GGGTGATGGT	TCACGTAGTG
1681	GGCCATCGCC	CTGATAGACG	GTTTTTCGCC	CTTTGACGTT	GGAGTCCACG	TTCTTTAATA	GTGGACTCTT
1751	GTTCCTAACT	GGAAACACAC	TCAACCTTAT	CTCGGTCTAT	TCCTTTGATTT	TATAAGGGAT	TTTGGCCGAT
1821	TCGGCCTATT	GGTTAAAAAA	TGAGCTGATT	TAACAAAAAT	TTAACGCGAA	TTTAAACAAA	ATATTAACGC
1891	TTCAATATT	GGTGGCATT	TTCGGGGAAA	TGTGCGCGGA	ACCCCTATTT	GTATTATTTT	CTAAATACAT
1961	TCAATATGT	ATCCGCTCAT	GAGACAATAA	CCCTGATATA	TGCTTCAATA	ATATTGAAAA	AGGAAGAGTA
2031	TGAGTATTCA	ACATTTCCGT	GTCCGCCCTTA	TTCCCTTTTTT	TGCGCATTTT	TGCTTCTCTG	TTTGTGCTCA
2101	CCAGAAACG	CTGGTGAAG	TAAAGATGC	TGAAGATCAG	TTGCGTCCAC	GAGTGGGTTA	CATCGAATG
2171	GATCAACA	CGGTAAGAT	CCTTGAGAGT	TTTCGCCCCG	AAGAAGCTTT	TCCAATGATG	AGCACITTTA
2241	AAGTCTGCT	ATGTGGCGCG	GTATTATCCC	GTATTGACGC	CGGGCAAGAG	CAACTCGGTC	GCCGCATACA
2311	CTATTCTCAG	AATGACTTGG	TTGAGTACTC	ACCAGTCACA	GAAAGCATC	TTACGGATGG	CATGACAGTA
2381	AGAGAATTAT	GCAGTGCTGC	CATAACCATG	AGTGATAACA	CTCGGCCAA	CTTACTTCTG	ACAAAGATCG
2451	GAGGACCGAA	GGAGCTAACC	GCCTTTTTCG	ACAACATGGG	GGATCATGTA	ACTCGCCTTG	ATCGTTGGGA
2521	ACCGGAGCTG	AATGAAGCCA	TACCAACGGA	CGAGCGTGAC	ACCACGATGC	CTGTAGCAAT	GGCAACAACG
2591	TTGGGCAAAC	TATTAACTGG	CGAACTACTT	ACTCTAGCTT	CCGGCAACA	ATTAATAGAC	TGGATGGAGG
2661	CGGATAAAGT	TGCAGGACCA	CTTCTGCGCT	CGGCCCTTCC	GCTGGGCTGG	TTTATGCTG	ATAAATCTGG

FIG. 33C

2731 AGCCGGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCAGTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
 2801 GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAC GAAATAGACA GATCGCTGAG ATAGGTGCCT
 2871 CACTGATTAA GCATTGGTAA CTGTGAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAACTTCA
 2941 TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGAT AATCTCATGA AATCAATCCC TTAACGTGAG
 3011 TTTTCGTTC ACTGAGCTC AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTCTGC
 3081 GCCTAATCTG CTGCTTGCAA ACAAAAAAC CACCGCTACC AGCGGTGGTT TGTTCGCGG ATCAAGAGCT
 3151 ACCAATCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCTT TCTAGTGTAG
 3221 CCGTAGTTAG GCCACCACTT CAAGAATCTT GTAGCACCAG GTACATACCT CGCTGTGTA ATCTGTGTAC
 3291 CAGTGGCTGC TGCCAGTGGC GATAAGTCTT GTCTTACCG GTTGGACTCA AGCAGATAGT TACCGBATAA
 3361 GCGCAGCGG TCGGCTGAA CGGGGGTTC GTGCACACG CCCAGCTGG GAGAAAGCG GACAGGTATC
 3431 CTGAGATACC TACAGCTGA GCTATGAGAA AGCGCCACG TTCCCGAAGG GAAAACGCC GATATCTT
 3501 CGGTAAGCGG CAGGCTCGA ACAGGAGAG GCACGAGGA GCTTCCAGG GCTCGTCAGG GGGCGGAGC
 3571 TAGTCTGTC GGTTCGCC ACCTCTGACT TGAGCGTCGA TTTTGTGAT TGGCCCTTT GCTCACATGT
 3641 CTATGGAAA ACGCCAGCAA CGCGCCCTT GATTCGTGG ATAACCTAT TACCCTTTT GAGTGAGCTG ATACCGCTCG
 3711 TCTTCTCTG GTTATCCCTT GATTCGTGG ATAACCTAT TACCCTTTT TGGCCCTTT CTGCGCTTTT GCTCACATGT
 3781 CCGCAGCGA ACGACCGAGC GCAGCGAGTC AGTGAGCGAG GAAGCGGAG TCCCGACTGG AAGCGGGCA
 3851 CCTCTCCCG CGGTTGGCC GATTCATTAA TGACAGTGG CACACAGGTT GGCACCCAG GCTTTACACT TTATGCTTCC
 3921 GTGAGCGCAA CGCAATTAAT GTGAGTTAG TCACTCATTA GGCACCCAG CACACAGGAA ACAGCTATGA CCATGATTAC
 3991 GGCTCGTATG TTGTGTGAA TTGTGAGCG ATAAACAATTT KpnI XhoI
 4061 GCCAAGCGG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG GCGCCCTCTC GAGGTCAATC
 4131 ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAC AAGGTAAGA TTACCTGGTC AAAAGTGAAA
 4201 ACATCAGTTA AAAGGTGGTA TAAGTAAAT ATCGGTAATA AAGGTGGCC CAAAGTGAAA TTTTACTCTT
 4271 TCTACTATTA TAAAATTTGA GGATGTTTG TCGGTACTTT GATACGTCTT TTTTGTATGA ATTGGTCTTT
 4341 AAGTTTATTC GCGATTTGGA AATGCATATC TGTATTTAG TCGGTTTAA AGTTCGTTGC TTTTGTAAAT
 4411 ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCAT ATGCTAATTT GACATAAATTT TTGAGAAAAA
 4481 TATATATTCA GCGAATTC CAAATGAACA ATAATAAGAT TAAAATAGCT TGCCCCCGTT GCAGCGATGG
 4551 GTATTTTTC TAGTAAATA AAAGATAAC TTAGACTCAA AACATTTACA AAAACAACC CTAAGTCTCT
 4621 AAAGCCCAA GTGCTATGCA CGATCCATAG CAAGCCGAG CCAACCCAAC CCAACCCAGT CCACCCAGT
 4691 GCAGCCCACT GGCAATAGT CTCCACCCCG GGCATCATCA CCGTGAAGTG TCCGCACCCAG CGCACGCTCT
 4761 GCAGCCCAA AAAAAAAG AAGAAAAA AAGAAAAA AAGAAAAA GTGGGTCCGG GTCGTGGGGG
 4831 CCGAAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCGCTCCCTC CGCTTCCAAA GAAACGCCCC

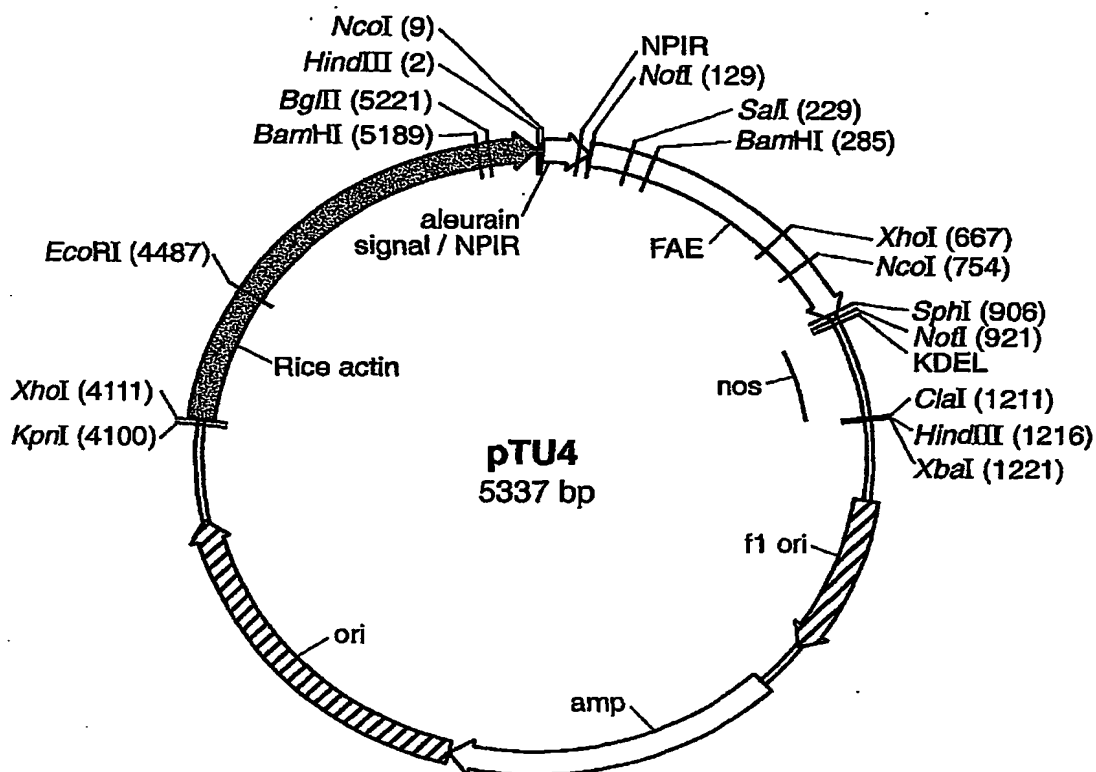
FIG.-33D

```

4901 CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCACACG TACCACCAAC ACCACCAACA
4971 CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCTCTCC GCCGCCGCC GTAACCAACC
5041 CGCCCCCTCTC CTCCTTCTTT CTCCTGTTTT TTTTTCGTC TTTTTCGAT CTTTGGCCTT GGTAGTTTGG
5111 GTGGGCGAGA GCGGCTTCGT CCCCCAGATC GGTGCGCGGG AGGGCGGGA TCTCGCGGT GGCCTCTCCG
                                     BamHI
                                     ~~~~~
5181 GCGGTGAGTC GGCCCCGATC CTCGCCGGGA ATGGGCTCT CGGATGTAGA TCTTCTTCT TTTTCTTTT
5251 TGTGGTAGAA TTGAATCCC TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA
5321 GCGTCTGCG GAGCTTTT GTAGC
                                     BglII
                                     ~~~~~

```

FIG._33E

**FIG. 34A**

```

      NgOI
      ~~~~~
HindIII
      ~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1  AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGCG TCGCGTGT GGCACGGCC GCGTGGCGG
      NotI
      ~~~~~
      . A S S S F A D S N P I R P V T D R A A A S T .
71  TCGCTCCTC CTCTCCTTC GCCGACTCCA ACCGATCG GCCGTCACC GACCGCGCG CCGCTCCAC
      . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
      SalI
      ~~~~~
      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTAACG
      BamHI
      ~~~~~
      . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGGACGAC AGCAGCAAG AAATAATCAC CGTCTTCCGT GGCACGTGTA GTGATACGAA
      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GATACTAAT ACACCTCTAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
      H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
      . Q Y P D Y A L T V T G H X L G A S L A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
      . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCAGCTG TCTGCGCAT ACGACAACAT CCGCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
      XhoI
      ~~~~~
      A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCTCGA GCCCAGATAC GACGCGATAT TTCCGGGTCA
      NgOI
      ~~~~~
      . H A N D G I P N L P P V E Q G Y A H G G V E Y .

```

FIG.-34B

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCAGTGGG GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
 . W S V D P Y S A Q N T F V C T G D E V Q C C E
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACCTGGG ATGAAGTGA GTGCTGTGAG
 SphI
 A Q G G Q G V N N A H T T Y F G M T S G A C T W
 841 GCCCAGGCG GACAGGGTGT GAATATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT
 NotI
 . P V A A A E P L K D E L *
 911 GGCGGTGCG GCGCGCGGAA CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAC ATTTGGCAAT
 981 AAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT
 1051 TAAGCATGTA ATAATTAA CA TGAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG
 1121 CAATTATACA TTTAATACGC GATAGAAAC AAAATATAGC GCGCAACTA GGATAAATTA TCGCGCGCGG
 HindIII
 ClaI XbaI
 1191 TGTATCTAT GTTACTAGAT CGATAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGAG
 1261 TCGTATFAC GCGCTCACT GCGCGTCTGT TTACAACGTC GTGACTGGGA AAACCTTGGC GTTACCCAAC
 1331 TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
 1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG CGCATTAAG CGCGCGCGGT
 1471 GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCTAGCGCC CGCTCTTTC GCTTCTTTC
 1541 CTTCCCTTCT TTACGGCACC TCGACCCCAA AAACTTGTAT TAGGTGATG GGGCTCCCTT TAGGGTTCCG
 1611 ATTTAGTGT CCGTTTTCG CCCTTTGACG TTGGAGTCCA CTTTCTTAA TAGTGAGCTC TTGTTCCAAA
 1681 CCTGATAGA CGGTTTTCG ACTCAACCTT ATCTCGGTCT ATTTTTCGA TTTATAAGG ATTTTCCGA TTTCCGCCCTA
 1751 CTGGACAAAC AATGAGCTGA TTTAACAAA ATTTAACGCG AATTTAACA AATATAATAC GCTTACAATT
 1821 TTGGTTAAA TTTTCGGGGA AATGTCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAATAT
 1891 TAGGTGGCAC GTGTCGCCCT TATTCCTTT TTTGCGGCGT TTTGCTTCC TGTTTTGT CACCCAGAAA
 1961 GTATCCGCTC ATGAGACAA ATCCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATTT
 2031 CAACATTTCC GTGTCGCCCT TATTCCTTT TTTGCGGCGT TTTGCTTCC TGTTTTGT CACCCAGAAA
 2101 CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA
 2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTCCTTCC TGAAGTCTG TAAAGTCTG
 2241 CATGTGGCG CGGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTTCT
 2311 AGAATGACTT GGTGAGTAC TCACCACTGA CAGAAAAGCA TCCTACGGAT GGCATGACAG TAAGAGATTT
 2381 ATGAGTGTCT GCCATAACCA TGAGTGATA CACTCGGCGC AACTTACTTC TGACAACGAT CGGAGGACCG

FIG.-34C

2451 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC
 2521 TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA
 2591 ACTATTAACT GCGGAATAC TTAATCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GCGCGATAAA
 2661 GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT CCGTTATTCG TGATAAATCT GGAGCCGGTG
 2731 AGCGTGGGTC TCBCGATATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCGATATCG TAGTTATCTA
 2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT
 2871 AAGCATGGGT AACTGTGAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACCTT CATTTTAAAT
 2941 TAAAAGGAT CTAGGTGAAG ATCCTTTTGG ATAATCTCAT GACCAAAATC CTTTAAACGTG AGTTTTCGTT
 3011 CCAGTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT CCAGCGGTGG TCTTGGAGATC CTTTTTTCCT GCGCGTAAATC
 3081 TGCTGCTTGC AAACAATAAA ACCACCGCTA CCGAGATACC GCCTACATAC CTTGCTGTGC TAAATCTGTT ACCAGTGGCT
 3151 TTTTTCGGA GGTAACTGGC TTCAAGAACT CTGTAGCAC AGCCAGCTT GGAGCGAAG ACCTACACCG AACTGAGATA
 3221 AGGCCACCCAC TTCAAGAACT CTGTAGCAC AGCCAGCTT GGAGCGAAG ACCTACACCG AACTGAGATA
 3291 GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGTGTGGACT CAAGACGATA CTTACCGGAT AAGCGCGCAGC
 3361 GGTGCGGCTG AACGGGGGT TCGTGACAC AGCCAGCTT GGAGCGAAG ACCTACACCG AACTGAGATA
 3431 CCTACAGCCT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAG GGGACAGGTA TCCGGTAAGC
 3501 GGCAGGGTCG GAACAGGAGA GCGCAGGAG GAGCTTCCAG GAGTATCCAG GGGGAAACGC CTGGTATCTT TATAGTCTCTG
 3571 TCGGTTCG CACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA TGGCTGCAT TTTCTTTTCTT GTTCTTTTCTT
 3641 AACGCCAGC ACGCGGCTT TTTTACGGTT CTTGCGCTTT ATTACCGCTT TTGATGAGC TGAATACCGCT CGCCGCGAGCC
 3711 GCGTTATCCC CCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAAC CGCTCTCTCC
 3781 GAACGACCGA CCGATTCATT AATGCACTG GCACGACAGG TTTCGCCGACT GGAAGCGGG CAGTGAGCGC
 3851 CCGCGGTTGG CCGATTCAAT ATGTGAGTTA GCTCACTCAT TAGGCACCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA
 3921 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA
 3991 TGTGTGTGG AATGTGAGC GGATAACAAT TTCACACAGG AACAGCTAT GACCATGATT ACGCCAAGCG

 4061 CGCAATTAAC CCTCACTAAA GGAACAABAA GCTGGGTACC GGGCCCCCCC TCGAGGTCT TCAATATGCTT
 4131 GAGAAAGAG TCGGGAATAGT CCAAAATAAA ACAAGGTAA GATTACCTGG TCAAAAGTGA AAACATCAGT
 4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCAAAAGTGA AATTTACTCT TTTCTACTAT
 4271 TATAAAATTT GAGGATGTTT TGTCCGGTACT TTGATACGTC ATTTTGTGAT GAATTTGGTT TTAAGTTTAT
 4341 TCGCGATTGG GAAATGCATA TCTGTATTGG AGTCGGTTT TAAAGTTCGTT GCTTTTGTAA ATACAGAGGG
 4411 ATTGTATATA GAAATATCTT TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT

 4481 CAGGCGAATT CCACAATGAA CAATAATAAG ATTAATAAG CTTGCCCCCG TTGCAGCGAT GGGTATTTT
 4551 TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAACATTTA CAAAACAC CCCTAAAGTC CTAAGGCCCA

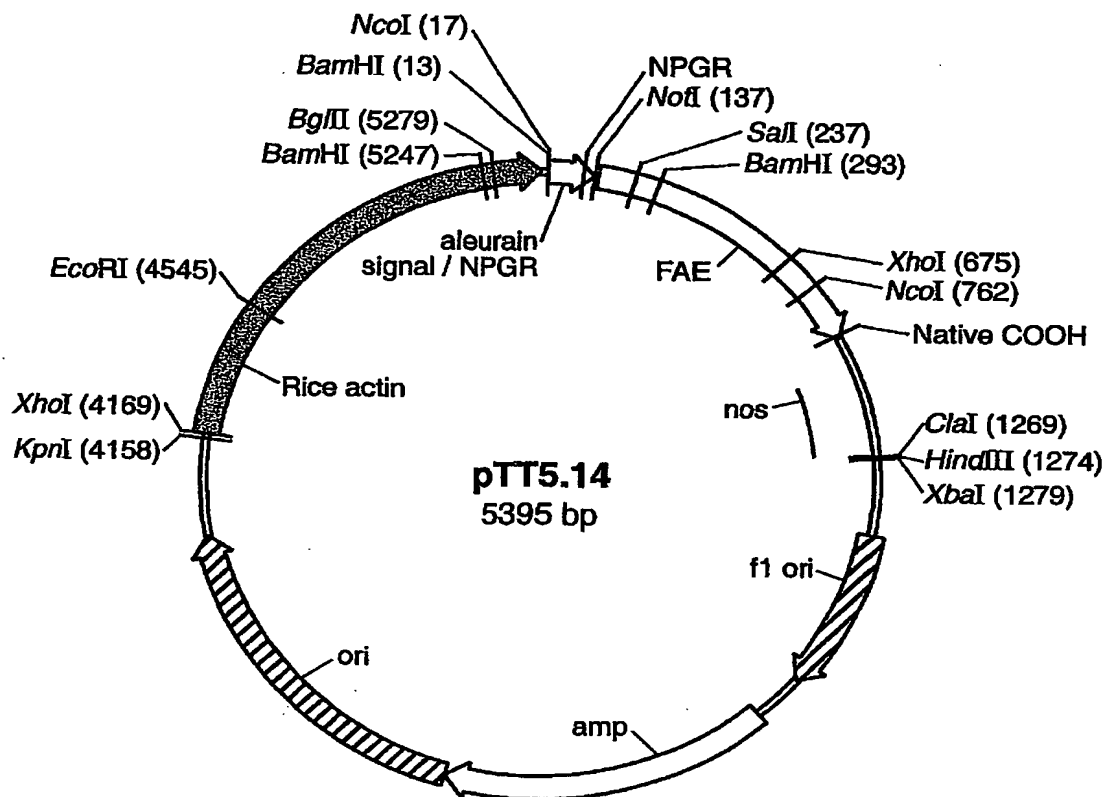
KpnI
 XhoI
 EcoRI

FIG.-34D

4621 AAGTGTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCACCCCA GTGCAGCCAA
4691 CTGGCAATA GTCTCCACCC CCGGCACCTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA
4761 AAAAAAAAAA AGAAGAAAA AAAAGAAAA GAAAAACAGC AGGTGGGTCC AGGTCTGTGGG GGCCGGAAAA
4831 GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCACA CCTAACACCA CCACCCACC CACCTCCTCC
4971 CCCCCTGCTG CCGGACGACG AGCTCCTCCC CCTCCCCCT CCGCCGCCGC CGGTAAACC CCGCCCCCTC
5041 TCCTCTTCTT TTCTCCGTTT TTTTTCGTT TCTGGTCTCG ATCTTGGCC TTGGTAGTTT GGGTGGGCGA
5111 GAGCGGCTTC GTCGCCAGA TCGGTCCCG GAGGGGCGG GATCTCGCG CTGGCTCTC CGGGCGTGAG
BamHI BglII

5181 TCGGCCCGGA TCCTCGCGG GAATGGGGCT CTCGGATGTA GATCTCTTT CTTCTTCTT TTTGTGTAG
5251 AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG
5321 CGGAGCTTTT TTGTAGC

FIG._34E

**FIG._35A**

```

      NgOI
      ~~~~~
      BamHI
      ~~~~~
1   M A H A R V L L L A L A V L A T A A .
    CCTGACGCG AGGATCCATG GCCACGCC GCCTCTCTCT CCTGGCGTC GCCGTGCTGG CCACGGCGCG
    NotI
71  . V A V A S S S F A D S N P G R P V T D R A A
    CGTCGCGTC GCCTCTCTCT CCTCTCTCG CCACTCCAAC CCGGCGCGC CCGTCACCGA CCGCGCGCGC
    NotI
    ..
141 A S T Q G I S E D L Y S R L V E M A T I S Q A A
     AGGTCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAT GCCCACTATC TCCCAAGCTG
     Sali
    ~~~~~
211 . Y A D L C N I P S T I I K G E K I Y N S Q T D .
     CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA GGGAGAGAAA ATTTACAATT CTCAAACTGA
     BamHI
    ~~~~~
281 . I N G W I L R D D S S K E I I T V F R G T G S
     CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATAATCACCG TCTTCCGTGG CACTGGTAGT
351 D T N L Q L D T N Y T L T P F D T L P Q C N G C
     GATACGAATC TACAACTCGA TACTAACTAC ACCCTCAGC CTTCGACAC CCTACCACAA TGCAACGGTT
    . E V H G G Y Y I G W V S V Q D Q V E S L V K Q .
421 GTGAAGTACA CCGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA GTCGAGTCGC TTGTCAAACA
    . Q V S Q Y P D Y A L T V T G H X L G A S L A A
491 GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCACKCCC TCGGCGCCTC CCTGGCGGCA
    L T A A Q L S A T Y D N I R L Y T F G E P R S G
561 CTCACGTGCG CCCAGCTGTC TCGGACATAC GACAACATCC GCCTGTACAC CTTCGGCGAA CCGCGCAGCG
    XhoI
    ~~~~~
631 . N Q A F A S Y M N D A F Q A S S P D T T Q Y F .
     GCAATCAGGC CTTGCGGTGG TACATGAACG ATGCCTTCCA AGCCTCGAGC CCAGATACGA CGCAGTATTT
    NgOI
    ~~~~~
    . R V T H A N D G I P N L P P V E Q G Y A H G G

```

FIG._35B

701 CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCC CCGGTGGAGC AGGGGTACGC CCATGGCGGT
 V E Y W S V D P Y S A Q N T F V C T G D E V Q C
 771 GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACCA CATTTGTCTG CACTGGGGAT GAAGTGCAGT
 . C E A Q G G Q G V N A H T T Y F G M T S G A .
 841 GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT TTTGGGATGA CGAGCGGAGC
 . C T W *
 911 CTGTACATGG TGATCAGTCA TTTTCAGCCTC CCGAGTGTA CCAGGAAGA TGGATGTCTT GGAGAGGGGG
 981 CCGGTAAACC ACTGAAGGAT GAGCTGTAAA GAAGCAGATC GTTCAAAACAT TTGGCAATAA AGTTTCTTAA
 1051 GATTGAATCC TGTGCGCGGT CTTCGATGTA TTATCATATA ATTCTGTGTG AATTACGTTA AGCATGTAT
 1121 AATTAAACATG TAATGCATGA CGTTATTAT GAGATGGGT TTTATGATTA GAGTCCCGCA ATTATACATT
 1191 TAATACCGGA TAGAAACAA AATATAGCGC GCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT
 HindIII
 ~~~~~  
 ClaI XbaI  
 ~~~~~  
 1261 TACTAGATCG ATAAGCTTCT AGAGCGGCCG GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGGC
 1331 CGCTCACATGG CCGTCGTTT ACACCGTCTG GACTGGGAA ACCCTGGCGT TACCCAACTT AATCGCCTTG
 1401 CAGCACATCC CCTTTTCGCC AGCTGGCGTA CTAGCGAGA GGGCCGCACC GATCGCCCTT CCCAACAGTT
 1471 GCGCAGCCTG AATGGCGAAT GGGACGCGCC CTAGCGGCCG GCATTAAAGC GCGCGGGTGT GGTGGTTACG
 1541 CGCAGCGTGA CCGCTACACT TGGCAGCGCC CTAGCGGCCG CTCTTTTCGC TTTCTTCCCT TCCCTTCTCG
 1611 CCACGTTGCG CGGCTTTCCC CGTCAAGCTC TAAATCGGG GCTTCCGAT GGGTTCCGAT TTAGTGTCTT
 1681 ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG GGGCATCGCC CTGATAGACG
 1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTTCTTAAATA GTGGACTCTT GTTCCAACT GGAACAACAC
 1821 TCAACCCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTCGCCGAT TCGGCCCTAT GGTAAAAA
 1891 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAAACAA ATATTAAACG TTACAATTTA GGTGGCACTT
 1961 TTCGGGGAAA TGTGCGCGGA ACCCTTATTT GTTTATTTT CTAAATACAT TCAATATATG ATCCGCTCAT
 2031 GAGACAATAA CCTGATATA TGTTCATAA ATATTGAAA AGGAAGAGTA TGAATATTCA ACATTTCCGT
 2101 GTCGCCCTTA TTCCCTTTT TCGGGCATTT TGCCTTCCCT CATCGAATCA CCCAGAAACG CTGGTGAAG
 2171 TAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA TCCAAATGAT AGCTCTCAACA GCGGTAAAGT
 2241 CCTTGAGAGT TTTTCGCCCG AAGAACGTTT TCCAAATGAT AGCTCTCAACA GAGTCTCAACA GCGGTAAAGT
 2311 GTATTATCCC GTATTGACGC CCGGCAAGAG CAACTCGCT CCGCATACA CTATTTCTCAG AATGACTTGG
 2381 TTGAGTACTC ACCAGTCACA GAAAGCATC TTACGGATGG CATGACAGTA AGAGATTAT GCAGTGTGCG
 2451 CATAAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAACCC
 2521 GCTTTTTTTC ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA
 2591 TACCAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG TTGCGCAAC TATTAACTGG
 2661 CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA

FIG.-35C

2731 CTTCTGGCGT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC
 2801 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG
 2871 TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA
 2941 CTGTGAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAAACCTCA TTTTAAATTT AAAGGATCT
 3011 AGGTGAAGAT CCTTTTGGAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTCGTTCC ACTGAGCGTC
 3081 AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTCTGTC GCGTAATCTG CTGCTTGCAA
 3151 ACAAAAAC CACCGTACC AGCGGTGGTT TGTTGGCCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG
 3221 TAATGGCTT CAGCAGAGCG CAGATACCA ATACTGTCTT TCTAGTGTAG CCGTAGTTAG CCCACCCTT
 3291 CAAAGACTCT GTAGCACCGC GTACATACCT CGTCTGTCTA ATCCTGTAC CAGTGGCTGC TGCCAGTGGC
 3361 GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA CCGCGAGCGG TCGGGGCTGA
 3431 CCGGGGGTTC GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA
 3501 GCTATGAGAA AGCGCCACGC TTCCCGAAG GAGAAAGCG GACAGGTATC CCGTAAGCGG CAGGGTCGGA
 3571 ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAACGCTT GGTATCTTTA TAGTCTGTTC GGGTTTCGCC
 3641 ACCTCTGACT TGAGCGTCTGA TTTTGTGTGAT GCTCGTCTAG GGGCGGAGC CTATGGAAAA ACGCCAGCAA
 3711 CCGGGCCTTT TTACGGTTCC TGGCCTTTTG GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCCGAGC
 3781 GATTCTGTGG ATAACCGTAT TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCCGAGC
 3851 GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAAACCG CCTCTCCCGG CCGGTTGGCC
 3921 GATTCATTAA TGCAGCTGGC ACGACAGGTT TCCCGAGTGG AAAGCGGCA GTGAGCGCAA CGCAATTAAT
 3991 GTGAGTTAGC TCACTCATTA GGCACCCGAG GCCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
 4061 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCGCG CAAATTAAACC

 4131 TCACTAAAGG GAACAAAAGC TGGGTACCGG GCGCCCTCTC GAGGTCAATC ATATGCTTGA GAAGAGAGTC
 4201 GGGATAGTCC AAAATAAAC AAAGGTAAAG TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAAGGTGGTA
 4271 TAAGTAAAAAT ATCGGTAAATA AAAGGTGGCC CAAAGTGAAA TTACTCTTT TCTACTATTA TAAAAATTGA
 4341 GGATGTTTTG TCGGTACTTT GATACGTCAT TTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTGGGA
 4411 AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCGTTGC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA

 4481 AATATCTTTA AAAAACCCAT ATGCTAATTT GACATAAATTT TTGAGAAAAA TATATATTCA GGCGAATTTCC
 4551 ACAATGAACA ATAAATAGAT TAAATAGCT TGCCCCCGTT GCAGCGATGG GTATTTTTC TAGTAAATA
 4621 AAAGATAAAC TTAGACTCAA AACATTTTACA AAAACAACCC CTAAAGTCTT AAAGCCCAA GTGCTATGCA
 4691 CGATCCATAG CAAGCCGAGC CCAACCCCAAC CCACCCCACT GCAGCCCACT GGCAAAATAGT
 4761 CTCACCCCCC GGCACATATCA CCGTGAGTGG TCCGACCCAC CGCAGCTCTC GCAGCCCAA AAAAAAAG
 4831 AAAGAAAAA AAGAAAAAGA AAAACAGCAG GTGGGTCCGG GTCTGTGGGG CCGGAAAAAGC GAGGAGGATC

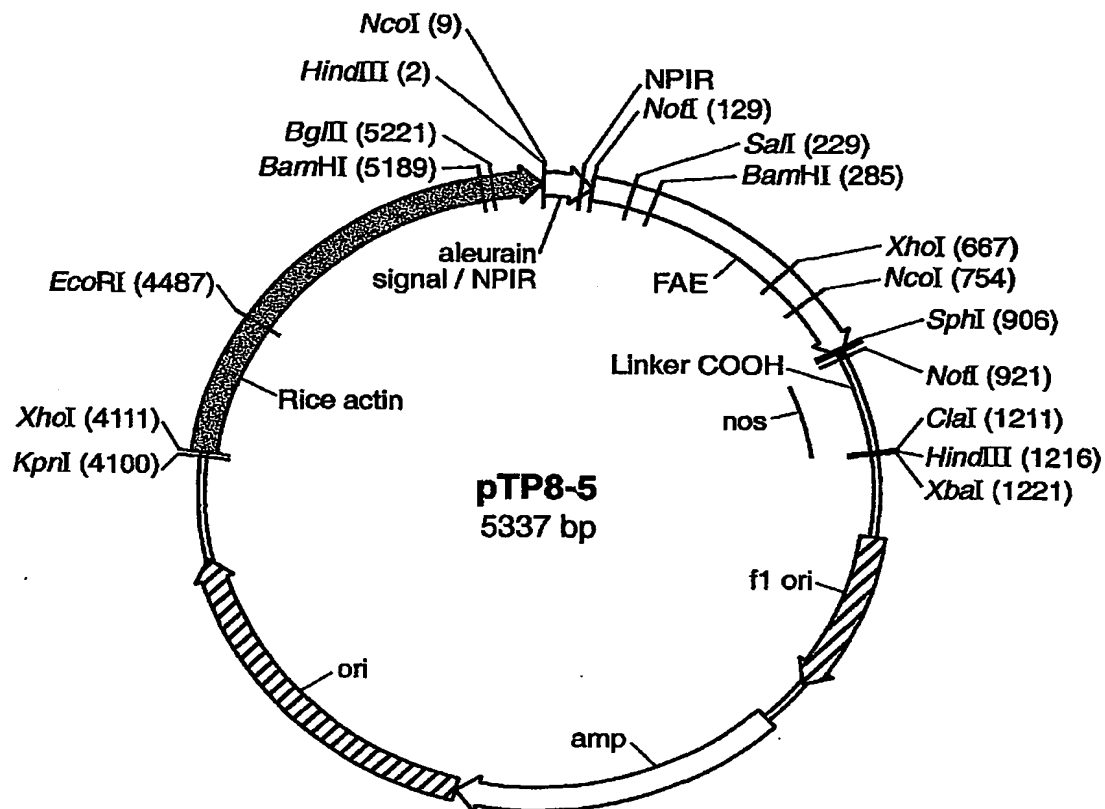
FIG.-35D

```

4901 GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAA GAAACGCCCC CCATCGCCAC TATATACATA
4971 CCCCCCCTC TCCTCCCATC CCCCACACCC TACCACACCC ACCACCCACCA CCTCCTCCCC CCTCGCTGCC
5041 GACGACGAG CTCCTCCCC CTCCTCCCTCC GCGGCCGCCG GTAACCAACC GCGCCCTCTC CTCCTTCTTT
5111 CTCCTGTTTT TTTTTCGCT CCGTCTCGAT CTTTGGCCTT GGTAGTTGG GTGGGCGAGA GCGGCTTCGT
BamHI
5181 CGCCACAGTC GGTGCGCGG AGGGCGGGA TCTCGCGGCT GCGCTCTCCG GCGGTGAGTC GGCCCGGATC
BamHI
~
5251 CTCGCGGGA ATGGGCTCT CGGATGTAGA TCTTCTTCT TTCTTCTTTT TGTGGTAGAA TTGAATCCC
5321 TCAGCATTTG TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA GCGTCTGCG GAGCTTTTTT
5391 GTAGC

```

FIG._35E

**FIG._36A**

```

NcoI
-----
HindIII
-----
      M A H A R V L L L A L A V L A T A A V A V
1  AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTGCGCG
      NotI
      . A S S S F A D S N P I R P V T D R A A A S T .
71  TCGCCTCCTC CTCTCCTTC GCGACTCCA ACCGATCCG GCCGTACCC GACCGCGCG CCGCCTCCAC
      . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141  GCAGGGCATC TCCGAAGACC TCTACAGCCG TTATAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
      Sall
      . . . . .
      D L C N I P S F I I K G E K I Y N S Q T D I N G
211  GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAACT GACATTAAACG
      BamHI
      . . . . .
      . W I L R D D S S K E I I T V F R G T G S D T N .
281  GATGATCCT CCGGACGAC AGCAGCAAG AAATAATCAC CGTCTCCGT GGCACGTGTA GTGATACGAA
      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351  TCTACAACTC GATACTAAT ACACCTCAC GCCTTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
      E G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421  CACGCTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGATC GCTTGTCAA CAGCAGGTTA
      . Q Y P D Y A L T V T G H X L G A S L A L T A .
491  GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCAGTGC
      . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561  CGCCGAGCTG TCTGGGACAT ACGACAACAT CGCCTGTATC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
      XhoI
      . . . . .
      A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631  GCCTTCCGCT CGTACATGAA CGATGCCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
      NcoI
      . H A N D G I P N L P P V E Q G Y A H G G V E Y .

```

FIG._36B

```

701 CTGATGCCAA CGACGGGATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCCATGGCG GTGTAGATA
    . W B V D P Y S A Q N T F V C T G D E V Q C C E
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACCTGGG ATGANGTGA GTGCTGTGAG
    sphI
    ~~~~~
    A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTITGGGAT GACGAGCGGC GCATGCACCT
    NotI
    ~~~~~
    . P V A A A *
911 GGCCGGTCCG GGCCGGGTAA CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT
981 AAAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGGGAT GATTATCATA TAAATTTCTGT TGAATTTACGT
1051 TAAGCATGTA ATAAATTAACA TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG
1121 CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG
    HindIII
    ~~~~~
    ClaI XbaI
    ~~~~~
1191 TGTATCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGCG CGGTGGAGCT CCAATTCGCC CTATAGTGAG
1261 TCGTATTACG CGGCTCACT GGCCGTCTGTT TTACAACTGC GTGACTGGGA AAACCTTGGC GTTACCCCAAC
1331 TTAATCGCCT TGCAGCACAT CCCCCTTTTC CCAGCTGCGG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG GCGCATTAAG CGCGCGCGGT
1471 GTGGTGGTTA CGCGCAGCGT GACCCTTACA CTGCGCAGCG CCTTAGCGCC CGCTCCTTTC GCTTTCCTTC
1541 CTTCCTTTCT CGCCACGTTT CCGGCTTTC TCGACCCCAA AAAACTTGTAT TAGGTGATG GTTCACGTAG TGGGCGCATCG
1611 ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGTAT TAGGTGATG GTTCACGTAG TGGGCGCATCG
1681 CCTGATAGA CGGTTTTTCG CCTTTTGACG TTGGAGTCCA CGTCTTTAA TAGTGGACTC TTGTTCCCAA
1751 CTGGAACAAC ACTCAACCTT ATCTCGGTCT ATTTTTCG ATTTTTCGGA ATTTTTCGGA TTTTCGGCCTA
1821 TGGTTAATAA AATGAGCTGA TTTAACAAA ATTTTTCG ATTTTTCGGA ATTTTTCGGA TTTTCGGCCTA
1891 TAGGTGGCAC TTTTCGGGGA AATGTGCGGA AACCTGATA AATGCTTCAA TAAATTTTAT TTTTTCGCTA
1961 GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA TAAATTTTAT TTTTTCGCTA TTTTCGGCCTA
2031 CAACATTTCC GTGTGCGCCT TATTTCCCTT TTTTCGGGCT ATTTTTCGGA AATGCTTCAA TTTTCGGCCTA
2101 CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA
2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTTCAATGA TGAGCACTTT TAAAGTTCTG
2241 CTATGTGGCG CGGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCGGCATA CACTATTCTC
2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT

```

FIG. 36C


```

2381 ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAAACGAT CGGAGGACC9
2451 AAGGAGCTAA CCGCTTTTTC GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAAACCGGAGC
2521 TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA
2591 ACTATTAACT GCGGAACCTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGATGGA GCGGATATAA
2661 GTTCAGGAGC CACTCTGTGG CTGCGGCTT CTGCGGCTT TGGGTATGCG TCGATAATCT GAGAGCCGGTG
2731 AGCGTGGGTC TCGCGGTATC ATTCGAGCAC TTGAGGATGA TCGATCGCTG AGATAGGTG CTCACATGATT
2801 CACGACGGGG AGTCAGGCAA CTATGGATGA CCAAGTTTAC TCAATATATAC TTTAGATTGA TTTAAACATT
2871 AAGCATGGT AACTGTCTAGA CCACTTTTTC ATATATCTAT GACCAAAATC CTTTAAACGTT CATTTTAAAT
2941 TTAAAGGAT CTAGGTGAAG ATCTTTTTC TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTAAACGTT AGTTTTCGTT
3011 CCACTGAGCG TCAGACCCCG AACAAAAA ACCACCGCTA CCAGCGGTGG TCTTGAGATC CTTTAAACGTT GCGCGTAATC
3081 TGCTGCTTGC AAACAACTGG TTCAGCAGAG CCGAGATACC AAATACTGTC TAATCTGTT TAACTCTGTT
3151 TTTTTCGAA GGTAACTGGC GTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCAGAT AAGCGCAGC
3221 AGGCCACCAC TTCAAGAACT CTGTAGCACC GGTGTGGACT CAAGACGATA GTTACCAGAT AAGCGCAGC
3291 GCTGCCAGTG GCGATAAGTC GGTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCAGAT AAGCGCAGC
3361 GGTGCGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT GGTGTGGACT CAAGACGATA GTTACCAGAT AAGCGCAGC
3431 CCTACAGCGT GAGCTATGAG AAAGCGCCAC GGTGTCCGAA GGTGTCCGAA GGTGTCCGAA GGTGTCCGAA
3501 GGCAGGGTGC GAACAGGAG GCGCAGGAG GAGCTTCCAG GGTGTCCGAA GGTGTCCGAA GGTGTCCGAA
3571 TCGGCTTTCG CCACCTCTGA CTTGAGCTGC GATTCTTGTG ATGCTCTGTA GGTGTCCGAA GGTGTCCGAA
3641 AACGCGCAGC AACGCGGCTT TTTTACGGTT CTTGAGCTGC GATTCTTGTG ATGCTCTGTA GGTGTCCGAA
3711 GCGTTATCCC CTGATTTCTGT GGATAACCGT ATTACCGCTT TCGTGGCTT TCGTGGCTT TCGTGGCTT
3781 GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAAC CGCTCTCCC
3851 CGCGCGTTGG CCGATTCAAT AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAGCGGG CAGTGAGCGC
3921 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTTACA CTTTATGCTT CCGGCTCGTA
3991 TGTGTGTGG AATTGTGAGC GGATAACAAT TTCACACACG AACACGCTAT GACCATGATT ACGCCAAGCG

                                KpnI
                                ~~~~~
4061 CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTCAAT TCATATGCTT
4131 GAGAGAGAG TCGGATAGT CCAAAATAAA ACNAAAGTAA GATTACCTGG TCAAAAAGTGA AAACATCAGT
4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCCAAAGTGA AATTTACTCT TTTCTACTAT
4271 TATAAAAATT GAGGATGTTT TGTGCGTACT TTGATACGTC ATTTTGTAT GAAATGGTTT TTAAGTTTAT
4341 TCGCGATTGG GAAATGCATA TCTGTATTGG AGTCGGTTTT TAAGTTCTGT GCTTTTGTAA ATACAGAGGG
4411 ATTTGTATAA GAAATATCTT TAAAAAACCC ATATGCTAAT TTAGACATAAT TTTTGAGAAA AATATATATT

                                EcoRI
                                ~~~~~
4481 CAGGCGAATT CCACAATGAA CAATAATAAG ATTAAAAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT

```

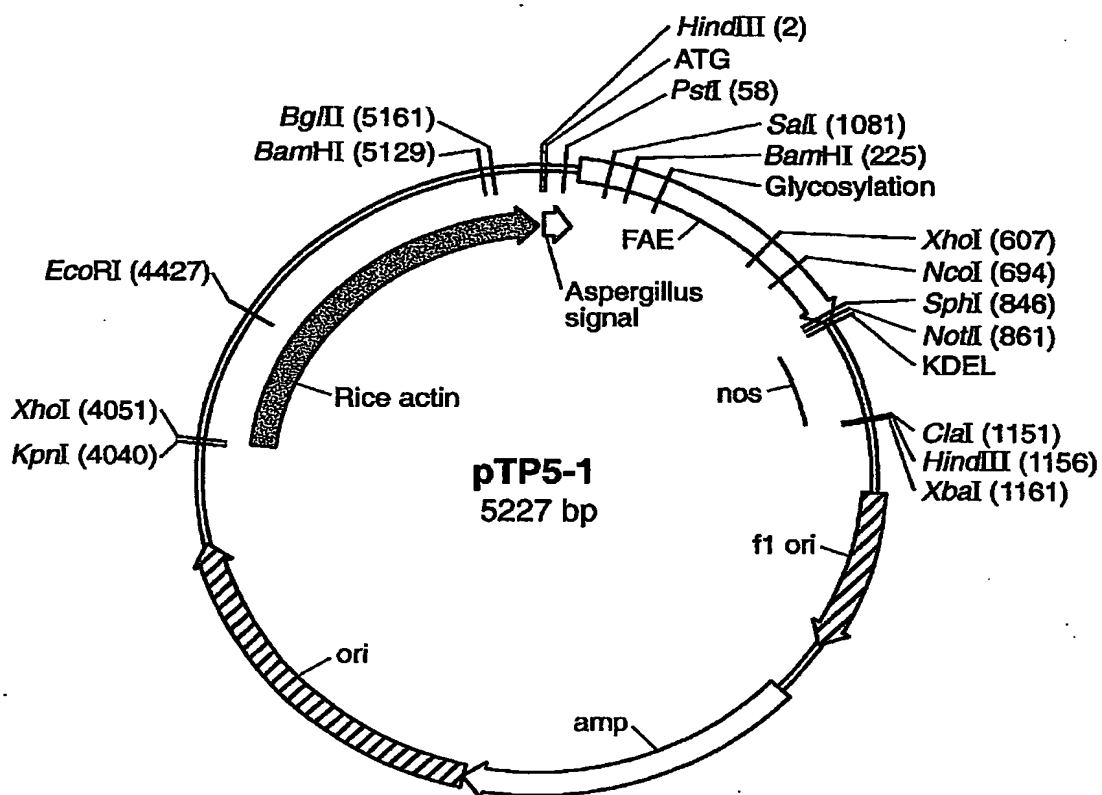
FIG._36D

```

4551 TCTAGTAAAA TAAAGATAA ACTTAGACTC AAACATTTA CAAAACAAC CCTAAAGTC CTAAGGCCA
4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCACCCCA GTGCAGCCAA
4691 CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA
4761 AAAAAAAA AGAAGAAAA TCGGAGCAG CGACGAGGC CGGCCCTCCC TCCGTTCCA AAGAAACGCC CCCCATCGCC
4831 GCGAGGAGGA TCGGAGCAG TCTCTCTCCA TCTCTCTCCA CCTACCACA CCACCACCAC CACCTCCTCC
4901 ACTATATACA TACCCCTCCC AGCTCTCTCC CCTCTCTCC CCGCGCGCG CGGTAACCA CCGGCCCTC
4971 CCCCTCGCTG CCGGACGACG TTTCTCTCTT TTTCTCTCTT CTGCGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA
5041 TCCCTCTTCT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
5111 GAGCGGCTTC GTCGCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCGCG CTGGCGTCTC CGGGCGTGA
          BamHI          BglII
          ~~~~~
5181 TCGGCCGGGA TCCTCGCGGG GAATGGGGCT CTCGGATGA GATCTCTTT CTTTCTTCTT TTTGTGGTAG
5251 AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTCTCTT TTCAATGATT GTGACAAATG CAGCCTCGTG
5321 CCGAGCTTTT TTGTAGC

```

FIG. 36E

**FIG._37A**

```

HindIII
~~~~~
      M K Q F S A K H V L A V V V T A G H A L A
      1 AAGCTTAACA TGAAGCAGTT CTCCGCCAA CAGTCTCTCG CAGTTGTGCT GACTGCAGGG CAGGCCTTAG
      . A S T Q G I S E D L Y S R L V E M A T I S Q A .
      71 CAGCCTCTAC GCAAGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC
      ~~~~~
      . A Y A D L C N I P S T I I K G E K I Y N S Q T
      141 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAACT
      ~~~~~
      BamHI
      ~~~~~
      D I N G W I L R D D S S K E I I T V F R G T G S
      211 GACATTAACG GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACCTGGTA
      . D T N L Q L D T N Y T L T P F D T L P Q C N G .
      281 GTGATACGAA TCTACAACATC GATACTAACT ACACCCCTAC GCCTTTGAC ACCCTACCAC AATGCAACGG
      . C E V H G G Y Y I G W V S V Q D Q V E S L V K
      351 TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA
      Q Q V S Q Y P D Y A L T V T G H X L G A S L A A
      421 CAGCAGGTTA GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACAC CCTCGGCGCC TCCCTGGCGG
      . L T A A Q L S A T Y D N I R L Y T F G E P R S .
      491 CACTCACTGC CGCCAGCTG TCTCGGACAT ACGACAACAT CCGCTGTATC ACCCTCGGCG AACCGCGCAG
      ~~~~~
      XhoI
      ~~~~~
      . G N Q A F A S Y M N D A F Q A S S P D T T Q Y
      561 CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCTCGA GCCCAGATAC GACGCAGTAT
      ~~~~~
      NcoI
      ~~~~~
      F R V T H A N D G I P N L P P V E Q G Y A H G G
      631 TTCCGGGTCA CTCATGCCAA CGACGGCATC CCAACCTGC CCCGGGTGGA GCAGGGGTAC GCCCATGGCG
      . V E Y W S V D P Y S A Q N T F V C T G D E V Q .
      701 GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACCTGGG ATGAAGTGCA
      . C C E A Q G G Q G V N N A H T T Y F G M T S G
      771 GTGCTGTGAG GCCCAGGCG GACAGGGTGT GAATAATGCG CACAGACTT ATTTTGGGAT GACGAGCGGC

```

FIG.-37B

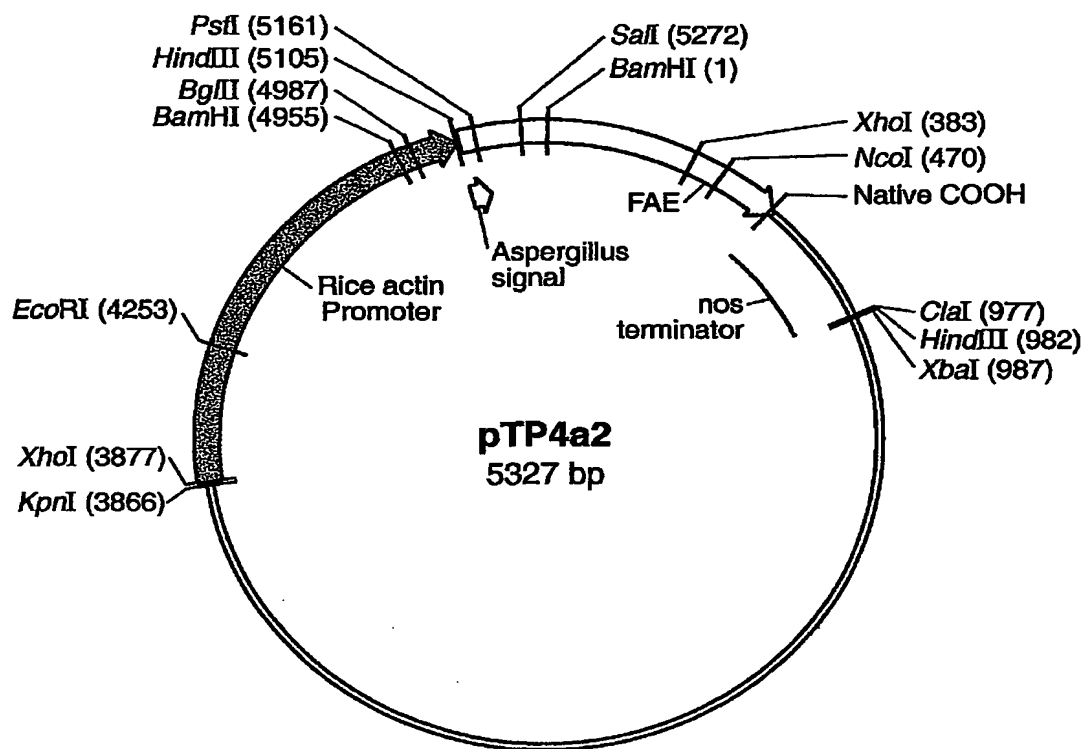
SphI									
NotI									
A	C	T	W	P	V	A	A	A	E
P	L	K	D	E	L	*			
841	GCATGCACCT	GGCCGGTCCG	GGCCGGCGAA	CCACTGAAG	ATGAGCTGTA	AAGAGCAGA	TCGTTCAAAC		
911	ATTGGCAAT	AAAGTTCTT	AGATTGAAT	CCTGTTGCCG	GTCTTGCGAT	GATTATCAT	TAATTTCTGT		
981	TGAATTACGT	TAAGCATGTA	ATAATTACA	TGTAATGCA	GACGTTATTT	ATGAGATGGG	TTTTTATGAT		
1051	TAGAGTCCCG	CAATTATACA	TTTAATACGC	GATAGAAAC	AAAATATAGC	GCACAACATA	GGATAAATTA		
HindIII									
ClaI									
XbaI									
1121	TCGCGCGCGG	TGTCATCTAT	GTACTAGAT	CGATAAGCTT	CTAGAGCGGC	CGGTGGAGCT	CCAATTGCGC		
1191	CTATAGTGAG	TCGTATTACG	CGCGTCACT	GGCCGTCGTT	TTACAACGTC	GTGACTGGGA	AAACCTGGC		
1261	GTATCCCAAC	TTAATCGCCT	TGCAGCACAT	CCCCCTTTCG	CCAGCTGGCG	TAATAGCGAA	GAGGCCCGCA		
1331	CCGATCGCCC	TTCCCAACAG	TTGCGCAGCC	TGAATGGCGA	ATGGGACGCG	CCCTGTAGCG	GCGCATTAG		
1401	CGCGGCGGGT	GTGGTGGTTA	CGCGCAGCGT	GACCGCTACA	CTTGCCAGCG	CCCTAGCGCC	CGCTCCTTTC		
1471	GCTTCTTCC	CTTCCCTTCT	CGCCACGTC	GCCGGCTTTC	CCCGTCAAGC	TCTAAATCGG	GGGCTCCCTT		
1541	TAGGGTTCCG	ATTAGTGCT	TTACGGCAC	TCGACCCCAA	AAAACTTGAT	TAGGGTGATG	GTTCACGTTAG		
1611	TGGGCCATCG	CCCTGATAGA	CGGTTTTCG	CCCTTTGACG	TTGGAGTCCA	CGTTCTTTAA	TAGTGGACTC		
1681	TTGTTCCTAA	CTGGACAAC	ACTCAACCT	ATCTGGTCT	ATTCCTTTGA	TTTATAAGGG	ATTTTGCCGA		
1751	TTTCGGCCCA	TTGGTTAAA	AATGAGCTGA	TTTAAACAAA	ATTTAACCGG	AAATTTAACA	AAATATTAC		
1821	GCTTACAAAT	TAGGTGGCAC	TTTTCGGGGA	AATGTGCGCG	GAACCCCTAT	TTGTTTATTT	TTCTAAATAC		
1891	ATTCAAAAT	GTATCCGCTC	ATGAGACAAT	AACCTTGATA	AATGCTTCAA	TAATATTGAA	AAAGGAAGAG		
1961	TATGAGTATT	CAACATTTC	GTGTCGCCCT	TATTCCTTTT	TTTGGGCGAT	TTTGCCTTCC	NGTTTTTGCT		
2031	CACCCAGAAA	CGCTGGTGAA	AGTAAAGAT	GCTGAAGATC	AGTTGGGTGC	ACGAGTGGGT	TACATCGAAC		
2101	TGGATCTCAA	CAGCGGTAAG	ATCCTTGAGA	GTTTTCGCC	CGAAGAACGT	TTTCCAATGA	TGAGCACTTT		
2171	TAAGTCTTG	CTATGTGGCG	CGGTATTATC	CCGTATTGAC	GCCGGSCAAG	AGCAACTCGG	TCGCCCCATA		
2241	CACATATCTC	AGATGACTTT	GGTTGAGTAC	TCACCAGTCA	CAGAAAAGCA	TCTTACGGAT	GGCATGACAG		
2311	TAAGAGATT	ATGCAGTGCT	GCCATAACCA	TGAGTGATAA	CACGTGGGCC	AACCTTACTTC	TGACAAACGAT		
2381	CGAGGACCG	AAGGAGCTAA	CCGCTTTTTC	GCACAACATG	GGGGATCATG	TAACCTCGCCT	TGATCGTTGG		
2451	GAACCGGAGC	TGAATGAAGC	CATACCACAA	GACGAGCGTG	ACACCACGAT	GCCTGTAGCA	ATGGCAACAA		
2521	CGTTGGGCAA	ACTATTAACT	GGCGAACTAC	TTACTCTAGC	TTCCGGGCAA	CAATTAATAG	ACTGGATGGA		
2591	GGCGGATAAA	GTTCAGGAC	CACCTCTGCG	CTCGGCCCTT	CGGGCTGGCT	GGTTTATTGC	TGATAAATCT		
2661	GGAGCCCGTG	AGCGTGGGTC	TCGCGGTATC	ATTGCAGCAC	TGGGGCCAGA	TGGTAAAGCC	TCCCCTATCG		

FIG.-37C

2731 TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAATAGA CAGATCGCTG AGATAGGTGC
 2801 CTCACCTGATT AAGCATGGT AACTGTGAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACIT
 2871 CAATTTTAAT TTAAGAAGGAT CTAGGTGAAG ATCCTTTTGG ATAACTCTCAT GACCAAAATC CCTTAAACGTG
 2941 AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAGAATCT TCTTGAGATC CTTTTTTTCT
 3011 GCCTGTAATC TGCTGCTTGC AAACAATAAA ACCACCGCTA CAGCGGTGG TTTGTTTGGC GGATCAAGAG
 3081 CTACCAACTC TTTTTCGGAA GGTAAGTGGC TTACGACAGG CGCAGATACC AATACTGTC CTTCFAGTGT
 3151 AGCCGTAGTT AGGCCACCAC TTCAAGAATC GGTCTTACC GGTTCGACT CAAGACGATA GTTACCCGAT
 3221 ACCAGTGGCT GCTGCCAGTG AACGGGGGT TCGTGCACAC AGCCAGCTT AGCGGAACG ACCTACACCG
 3291 AAGCGCAGC CCTACAGCGT GAGCTATGAG AAAGCGCCAC GACTTCCGAA GGGGAAACGC CGGACAGGTA
 3361 AACTGAGATA GGCAGGTCG GAACAGGAGA CACCTCTGA CTTGAGCTCA ATGCTCGTCA GGGGGCGCGA
 3431 TCCGCTAAGC TCAGGCTTCG AACGCCAGC TTTTACGGTT CTTGAGCTCA ATGCTCGTCA GGGGGCGCGA
 3501 TATAGTCTTG GCGGCTTCG AACGCCAGC TTTTACGGTT CTTGAGCTCA ATGCTCGTCA GGGGGCGCGA
 3571 GCTATGGA AAAGGCCAGC TTTTACGGTT CTTGAGCTCA ATGCTCGTCA GGGGGCGCGA
 3641 GTTCTTCTT CCGTTATCCC CTGATTCGT GGATAACCGT ATTACCGCTT TTGAGTGAGC TGATACCGCT
 3711 CGCCGACGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCA ATACGCAAC
 3781 CGCTCTCCC CGCGCTTGG CCGATTCAAT AATGACGCTG GCACGACAGG TTTCCCGACT GGAAGCGGG
 3851 CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT
 3921 CCGGCTCGTA TGTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG AACACGCTAT GACCATGATT
 3991 ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTCTAT
 4061 TCATATGCTT GAGAAGAGAG TCGGGATAGT CCAAAATAAA ACAAGGTAA GATTACCTGG TCAAAAGTGA
 4131 AAACATCAGT TAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAGGTGG CCCAAAGTGA AATTACTCT
 4201 TTTCTACTAT TATAAAATTT GAGGATGTTT TGTCGGTACT TTGATACGTC ATTTTGTAT GAAATTGGTTT
 4271 TTAAGTTTAT TCGCGATTGG GAAATGCATA TCTGTATTTG AGTCGGTTT TAAGTTCTGT TCTTTTGTAA
 4341 ATACAGAGGG ATTTGTATTA GAAATATCTT TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA
 4411 AATATATATT CAGGCGAATT CCACAATGAA CAATAATAAG ATTAATAAG CTGTCCCCCG TTGCAGCGAT
 4481 GGGTATTTT TCTAGTAAAA TAAAGATAA ACTTAGACTC AAAACATTTA CAAAACAAAC CCTTAAAGTC
 4551 CTAAAGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCAACCCA ACCCAACCCA
 4621 GTGCAGCCAA CTGGCAATA GTCTCCACC CCGGCACTAT CACCGTGAAT TGTCCGCACC ACCGCACGTC
 4691 TCGCAGCCAA AAAAAAAA AGAAGAAAA AAAAAACAGC AGGTGGGTCC GGGTCGTGGG

FIG. 37D

FIG. 37E

**FIG._38A**

BamHI

 . I L R D D S S K E I I T V F R G T G S D T N L
 1 GATCCTCCGC GACGACAGCA GCAAAGAAT AATCACCGTC TTCGGTGGCA CTGGTAGTGA TACGAATCTA
 Q L D T N Y T L T P F D T L P Q C N G C E V H G
 71 CAACTCGATA CTAACATACAC CCTCAGCCTT TCGACACCC TACACAATG CAACGGTTGT GAAGTACACG
 . G Y Y I G W V S V Q D Q V E S L V K Q Q V S Q .
 141 GTGGATATTA TATTGGATGG GTCTCCGTCC AGGACCAAGT CGAGTCGCTT GTCAACACAGC AGGTTAGCCA
 . Y P D Y A L T V T G H X L G A S L A A L T A A
 211 GTATCCGGAC TACGCGCTGA CCGTGACCGG CCACKCCCTC GCGCCTCCC TGGCGGCACT CACTGCCGCC
 Q L S A T Y D N I R L Y T F G E P R S G N Q A F
 281 CAGCTGCTCG CGACATACGA CAACATCCGC CTGTACACCT TCGGCGAACC GCGCAGCGGC AATCAGGCCT
 XhoI

 . A S Y M N D A F Q A S S P D T T Q Y F R V T H .
 351 TCGCGTCGTA CATGAACGAT GCCTTCCAAG CCTCGAGCCC AGATACGACG CAGTATTTC CCGTCACTCA
 NcoI

 . A N D G I P N L P P V E Q G Y A H G G V E Y W
 421 TGCCAAACGAC GGCATCCCAA ACCTGCCCCC GGTCGAGCAG GGTACGCC C ATGGCGGTGT AGAGTACTGG
 S V D P Y S A Q N T F V C T G D E V Q C C E A Q
 491 AGCGTTGATC CTTACAGCGC CCAGAACACA TTTGTCTGCA CTGGGGATGA AGTCAGTGC TGTGAGGCCC
 . G G Q G V N N A H T T Y F G M T S G A C T W * .
 561 AGGCGGACA GGTGTGAAT AATGCGACA CGACTTATTT TGGGATGACG AGCGGAGCCT GTACATGGTG
 . *
 631 ATCAGTCATT TCAGCCTCCC CGAGTGATCC AGGAAGATG GATGTCCTGG AGAGGGGGCC GCGTAACCCAC
 701 TGAAGGATGA GCTGTAAAGA AGCAGATCGT TCAACACATTT GGCATATAAG TTCTCTAAGA TTGAATCCTG
 771 TTGCCGGTCT TCGATGATTT ATCATATAAT TTCTGTGAA TTACGTTAAG CATGTAATAA TTAACATGTA
 841 ATGCATGACG TTATTATGA GATGGGTTTT TATGATTAGA GTCCCGCAAT TATACATTTA ATACGGGATA
 ClaI
 911 GAAACAAAA TATAGCGCGC AAAC TAGGAT AAATTATCGC GCGCGGTGTC ATCTATGTTA CTAGATCGAT
 XbaI

 HindIII

FIG.-38B

981 AAGCTTCTAG AGCGCCGGT GGAGCTCCAA TTCGCCCTAT AGTGATCGT ATACGCGG CTCACCTGGCC
 1051 GTCGTTTAC AACGTCGTGA CTGGGAAAC CTGGCGTTA CCAACTTAA TCGCCTTGCA GCACATCCCC
 1121 CTTTCGCCAG CTGGCGTAAT AGCGAAGAG CCGCACCGA TCGCCTTCC CAACAGTTGC GCAGCCTGAA
 1191 TGGCGAATGG GACGCGCCCT GTAGCGGCG ATTAAGCGG ATTAAGCGG GCGGGTGTG GCGGTGAGC
 1261 GCTACACTTG CACGCGCCCT AGCGCCCGCT CTTTCGCTT CTTTCGCTT TCTTCCTTC CTTTCGCTC
 1331 GCTTCCCGG TCAAGCTCTA AATCGGGGG TCCCTTAGG TCCCTTAGG GTTCGGATTT AGTGCTTTAC
 1401 CCCCCAAA CTTGATAGG GTGATGGTTC AGGTAGTGG CCAATCGCCTT GATAGACGGT TTTTCGCCCT
 1471 TTGACGTTGG AGTCCACGTT CTTTAAATAG GCACTCTTGT TCCAAACTG AACCAACTC TTTTCGCCCT
 1541 CCGTCTATTC TTTTGAATTA TAAGGATTT TGCCGATTT TCGCCTTAGG GGCCTATTG GGCCTATTG
 1611 ACATAAATTT AACGCGAAT TTAACAAAAT ATTAACGCTT ACAAATTAGG TGGCACTTTT CCGGAAATG
 1681 TGGCGGGAAC CCTATTGTT TATTTTTCT AAATACATC AAATATGTAT CCGCTCATG CCGCTCATG
 1751 CTGATAAATG CTTCAATAAT ATTGAAAAG GAAGATATG GAAGATATG AGTATTCAC ATTTCCGTG
 1821 CCGTTTCTTG CCGCATTTTG CTTTCTGTT TTTTCTGCT TTTTCTGCT CAGAAACGCT GGTGAAAGT
 1891 AAGATCAGTT GGTGACAGA GTGGGTIACA TCGAACTGGA TCTCAACAGC GGTAAAGATCC TTTGAGAGTT
 1961 TCGCCCGAA GAACGTTTTC CAATGATGAG CACTTTTAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT
 2031 ATTGACGCC GGCAGAGCA ACTCGGTGCG CGCATACAT ATTCTCAGAA TACTTTCAGAA TACTTTCAG
 2101 CAGTACAGA AAAGCATCTT ACGGATGGCA TGACAGTAAG AGAATATATG AGATATGCA ATTTCCGTG
 2171 TGATAACACT CCGGCCAAT TACTTCTGAC AACGATCGA ACGATATG GACCGAAGG AGTAAACCG
 2241 AACATGGGG ATCATGTAAC TCGCCTTGAT CCGTGGGAC CGGAGCTGA TGAAGCCATA CCAACCGAC
 2311 AGCGTACAC CACGATGCC TTAGCAATGG CAACAACGTT GCGCAAACTA TTAACCTGGC AACTACTTAC
 2381 TCTAGCTTCC CCGCAACAAT TAATAGACTG GATGAGGCG GATGAGGTTG CAGGACCACT TCTGCGCTC
 2451 GCCCTTCCG CTGGCTGGTT TATTGCTGAT AAATCTGGG GTATCTGAG CCGGTGAGC TGGGTCTCG
 2521 CAGCACTGG GCCAGATGGT AAGCCCTCC GTATCTGAT TATCTACAG ACGGGAGTC AGGCAACTAT
 2591 GGATGAACGA AATAGACAGA TCGCTGAGT AGGTGCTCA CTGATTAAGC TTTAATTAAGC GTGAAGATC
 2661 GTTACTCAT ATATACTTTA GATTGATTA AAATCTCAT TTTAATTAAGC TTTAATTAAGC GTGAAGATC
 2731 TTTTGTATAA TCTCATGACC AAATCCCTT AAGTGAATTT AAGTGAATTT TCTGTTCCAC ACCCGTACA
 2801 AAAGATCAA GGATCTTCTT GAGATCCCTT TTTTCTGCG CAACTCTTTT TCCGAAGGTA ACTGGCTTCA
 2871 CCGCTACCAG CCGTGGTTTG TTTGCCGGAT CAAGAGCTAC TAGTGTAGC GTAGTGTAGC CCACTTCA
 2941 GCAGAGCGCA GATACCAAAT ACTGTCTTTC CTTGTGCTA CTTGTGCTA GTGGCTGCT GCGGTGAGC
 3011 AGCACCGCTT TGGACTCAAG ACGATAGTTA CCGGATAAG CCGGATAAG CCGGATAAG CCGGATAAG
 3081 CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAG CCGGATAAG CCGGATAAG CCGGATAAG
 3151 GCACACAGCC CAGCTTGGAG CCAACGACCT ACACCGAAT CAGGATATC GTAAAGCGGCA GTAAAGCGG
 3221 CGCCACGCTT TCCAGGGG AAACGCTGG TATCTTATA TATCTTATA GTCTGTCTG GTCTGTCTG
 3291 ACAGGGAGC TTTTGTGATG TCGTCAGGG GCGCGAGGCT ATGGAATAAC GCGCGAGCA GCGCGAGCA
 3361 AGCGTCAAT TTTTGTGATG TCGTCAGGG GCGCGAGGCT ATGGAATAAC GCGCGAGCA GCGCGAGCA

FIG.-38C

```

3431 ACGGTTCTTG GCCTTTTGCT GGCCTTTTGC TCACATGTTT TTTCTCTGCGT TATCCCTCTGA TTCTGTGGAT
3501 AACCGTATTA CCGCCTTTGA GTGAGCTGAT ACCGCTTCGC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG
3571 TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAACCCGCC TCTCCCCCGG CGTTGGCCGA TTCAATTAATG
3641 CAGCTGGCAC GACAGGTTTC CCGACTGGAA AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTAGCTC
3711 ACTCATTAGG CACCCAGGC TTACACATTT ATGCTTCGG CTGCTATGTT GTGTGGAATT GTGAGCGGAT
3781 AACAAATTCA CACAGGAAC AGCTATGACC AGTATTACGC CAAGCGCGCA ATTAACCCCTC ACTAAAGGGA

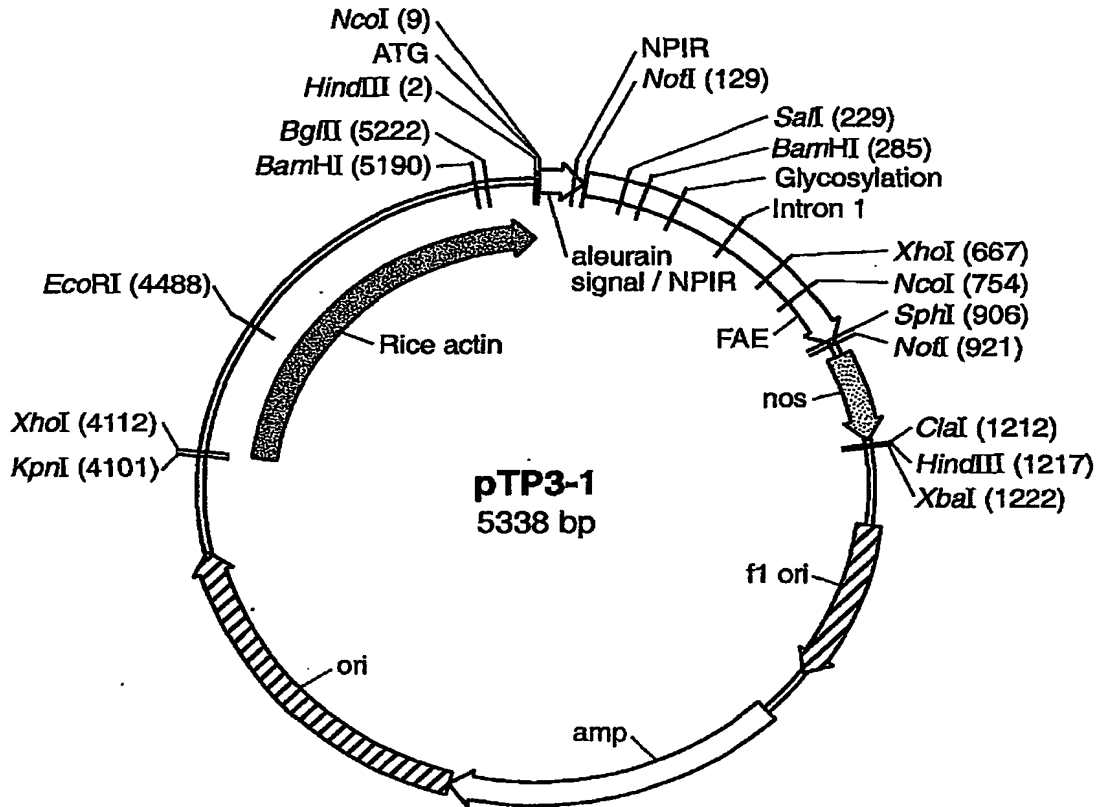
          KpnI          XhoI
          ~~~~~
3851 AAAAAAGCTG GGTACCGGGC CCCCCCTCGA GGTCAATTCAT ATGCTTGAGA AGAGAGTCGG GATAGTCCAA
3921 AATAAAACAA AGGTAAGATT ACCTGGTCAA AAGTGAAAC ATCAGTTAAA AGGTGGTATA AGTAAATAT
3991 CGGTAATAAA AGGTGGCCCA AAGTGAAATT TACTCTTTC TACTATTATA AAAATTGAGG ATGTTTGTTC
4061 GGTACTTTGA TACGTCAATTT TTGTATGAAT TGGTTTTTAA GTTTATTCGC GATTTGGAAA TGCATATCTG
4131 TATTGAGTC GGTTTTAAAG TTCGTTTGCTT TTGTAATAAC AGAGGGATTT GTATAAGAAA TATCTTTTAA
          EcoRI
          ~~~~~
4201 AAACCCATAT GCTAATTTGA CATAATTTT GAGAAATAA TATATTCAGG CGAATTCAC ACATGACAT
4271 AATAAGATTA AATAGCTTG CCCCCTTGC AGCGATGGGT ATTTTTCCTA GTAAATATAA AGATAAATTT
4341 AGACTCAAAA CATTTACAAA AACCAACCC AACCAACCC ACCCACTGC AGCCAAGT GCTATGCACG ATCCCTAGCA
4411 AGCCAGCCC AACCACACC AACCACACC CGCACCACCG CACGTCGCG AGCCAAATAA AGAATAAATA
4481 CACTATCACG GTGAGTTGTC GTGAGTTGTC GGTCCCGGT CGTGGGGCC GAAAGAGCGA GGAGGATCGC GAGCAGCGAC
4551 GAAAAAGAAA AACAGCAGGT AACAGCAGGT CTTCCAAAGA AACGCCCCC ATCGCCACTA TATACATACC CCCCCCTCTC
4621 GAGGCCCGGC CCTCCCTCCG CCAACCCCTA CCACACCCAC CACCACCCAC TCCTCCCCC FCGCTGCCGG ACGACGAGCT
4691 CTCCCATCCC CCTCCCCCT CCTCCCCCT CGCCGCCGGT AACCAACCCG CCCCCTCTCT CTTCTCTTCT CCGTTTTTFT
4761 TTTCTGTCCTG GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCT CCCAGATCGG
4831
          BamHI
          ~~~~~
4901 TGCGCGGGAG GGGCGGGATC TCGCGGCTGG CGTCTCCGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT
          BglII
          ~~~~~
4971 GGGGCTCTCG GATGTAGATC TTCTTTTCTT TTCTTTTCTT CTTCTTTTCT TGGTAGAATT TGAATCCCTC AGCATTTGTT
          HindIII
          ~~~~~
5041 ATCGGTAGTT TTTCTTTTCA TGATTGTGA CAAATGCAG CTCGTGCGGA GCTTTTTTGT AGCAAGCTTA

```

FIG. 38D

PstI
 M K Q F S A K H V L A V V V T A G H A L A A S .
 5111 ACATGAAGCA GTTCTCCGCC AACACGTC TCGCAGTTGT GGTGACTGCA GGGCAGCCT TAGCAGCCTC
 . T Q G I S E D L Y S R L V E M A T I S Q A A Y
 5181 TAGCAAGGC ATCTCCGAG ACCCTCTACG CCGTTAGTC GAAATGGCCA CTATCTCCA AGCTGCTAC
 Sali
 A D L C W I P S T I I K G E K I Y N S Q T D I N
 5251 GCCGACCTGT GCAACATTCC GTCGACTATT ATCAAGGAG AGAAATTTA CAATCTCAA ACTGACATTA
 B
 . G W
 5321 ACGGATG

FIG.-38E

**FIG. 39A**

NcoI
 ~~~~~  
 HindIII  
 ~~~~~  
 1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGGC TGGCGTGT GGCACGGCC GCGTGGCG
 W A H A R V L L L A L A V L A T A A V A V
 . A S S S F A D S N P I R P V T D R A A A S T .
 71 TGGCTCCTC CTCTCCTC GCGACTCA ACCGATCG GCGGTACC GACGGGCG GCGCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
 141 GCAGGGCATC TCGAAGACC TCTACAGCG TTAGTCGAA ATGCCACTA TCTCCCAAG TGCCTACGC
 Sali
 ~~~~~  
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTACG  
 BamHI  
 ~~~~~  
 281 GATGGATCCT CCGGACGAC AGCAGCAAG AATAATCAC CGTCTCCGT GGCACCTGTA GTGATACGA
 Glycosylation
 ~~~~~  
 351 TCTACAACTC GATACTAAT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACG TTGTGAAGTA  
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
 421 CACGGTGGAT ATTATATGG ATGGTCTCC GTCCAGACC AAGTCGATC GCTTGTCAA CAGCAGGTA  
 . Q Y P D Y A L T V T G H X L G A S L A A L T A .  
 491 GCCAGTATCC GGAATACGG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC  
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q  
 561 CGCCAGCTG TCTGGACAT ACGACAACAT CCGCTGTAC ACCTCGGCG AACCGCGCAG CGGCAATCAG

FIG. 39B

XhoI  
 A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
 631 GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA  
 . H A N D G I P N L P P V E Q G Y A H G G V E Y .  
 701 CTCATGCCAA CGACGGCATC CCAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCTTACA GCGCCCAAGAA CACATTGTTC TGCACCTGGG ATGAAGTGA GTGCTGTGAG  
 SphI  
 A Q G G Q G V N N A H T T Y F G M T S G A C T W  
 841 CCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT  
 NotI  
 KDEL  
 . P V A A A E T T E G \*  
 911 GCGCGGTGCG GCGCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTTGGCAA  
 981 TAAAGTTTCT TAAGATTGAA TCCTGTGTGCC GGTCTTGGCA TGATTATCAT ATAATTCTG TTGAAATTACG  
 1051 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTATTT TATGAGATGG GTTTTATGA TTAGAGTCCC  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAAACT AGGATAAATT ATCGCGCGCG  
 HindIII  
 ClaI  
 KbaI  
 1191 GTGTCACTTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTGCG CCTATAGTGA  
 1261 GTCGTATTAC GCGCGCTCAC TGGCCGTGCT TTACAAACGT CGTGAAGTGG AAAACCCCTGG CGTTACCCAA  
 1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
 1401 CTTCCCAACA GTTGCAGCAG CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGCGCGG

FIG. 39C

1471 TGTGGTGGTT ACGGCAGCG TGACCGCTAC ACTTGGCAGC GCCCTAGCG CCGCTCCCTT CGCTTCTCTC  
 1541 CCTTCCCTTC TCGCCACGTT CGCCGGCTTT CCGGCTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
 1611 GATTAGTGC TTACGGGAC CTCGACCCCA AAAAATGTA TTAGGTGAT GGTTCACGTA GTGGGCCATC  
 1681 GCCCTGATAG ACGGTTTTC GCGCTTTGAC GCGCTTTGAC GTTGGATCC ACCTTCTTAA ATAGTGGACT CTTGTTCCAA  
 1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTCTTTG TATTCTTTTG ATTATAAGG GATTTTGGCG ATTTCCGGCT  
 1821 ATTGGTTAAA AATGAGCTG AATTAACAA AATTAACGC GAATTTTAA CAAATATTA CCGTTACAA  
 1891 TTAGGTGGCA CTTTTCGGG AATGTCGGC GAATCCCTTA TTGTTTATT TTTCTAAATA CATTCAATA  
 1961 TGTATCCGCT CATGAGACAA TAACCTGTGAT AATGCTTCA ATATATTGA AAAAGGAAGA GTATGAGTAT  
 2031 TCAACATTTT CGTGTGCCC TTATTTCCCTT TGCTGAAGAT TTTTGGGCA TTTTGCCTTC TTACATCGAA TCACCCAGAA  
 2101 ACGTGGGTAA AAGTAAAGA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT  
 2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT  
 2241 GCTATGTGGC GCGGTATTAT CCGGTATTGA CCGGTATTGA CCGGTATTGA GAGCAACTCG GTCCCGCAT ACCTATTTCT  
 2311 CAGAAATGACT TGGTTGAGTA CTCACCACTC ATGAGTGATA ACCTGCGGC CAACCTACTT CTGACAACGA TCGGAGGACC  
 2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACCTGCGGC CAACCTACTT CTGACAACGA TCGGAGGACC  
 2451 GAAGGAGCTA ACCGCTTTT TGCACCAACT GCGGATCAT GTAACTCGCT TTGATCGTTG GGAACCGGAG  
 2521 CTGAATGAAG CCATACCAAA CGACGAGCGT GACACCAAG TGCCTGTAGC AATGGCAACA AGCTTGGCA  
 2591 AACTATTAA TGGCGAATA CTTACTCTAG CTTCCGGCA ACAATTAATA GACTGGATGG AGCGGATTA  
 2661 AGTTTCAGGA CCCTTCTGCG GCTCGGCTT CATTCGAGCA CTGGGGCCAG ATGGTAAAGC CTCCCGTATC TGGAGCCGGT  
 2731 GAGCGTGGGT CTCGCGGTAT GAGTCAGGCA ACTATGGATG AATGAAATAG ACAGATCGCT GAGTAGGTG CCTCACTGAT  
 2801 ACACGACGGG TAAGCATTTG TCTAGGTGAA GATCCCTTTT GATAATCTCA TGACCAAAAT CTTTAGATTG ATTTAAACT GAGTTTCTCT  
 2871 TTTAAAAGGA TCCACTGAGC GTACAGACCC CAAACCAAAA AACCACCGCT ACCAGCGGT TTTCTGAGAT CCGTTTCTTC TCGCGCTAAT  
 2941 TCCACTGAGC GTACAGACCC CAAACCAAAA AACCACCGCT ACCAGCGGT TTTCTGAGAT CCGTTTCTTC TCGCGCTAAT  
 3011 TCCACTGAGC GTACAGACCC CAAACCAAAA AACCACCGCT ACCAGCGGT TTTCTGAGAT CCGTTTCTTC TCGCGCTAAT  
 3081 TCCACTGAGC GTACAGACCC CAAACCAAAA AACCACCGCT ACCAGCGGT TTTCTGAGAT CCGTTTCTTC TCGCGCTAAT  
 3151 TCCACTGAGC GTACAGACCC CAAACCAAAA AACCACCGCT ACCAGCGGT TTTCTGAGAT CCGTTTCTTC TCGCGCTAAT  
 3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTTACATA CCTCGCTCTG TCAAGACGAT AGTTACCGGA TAAGGCGCAG  
 3291 TGTGCCCAGT GCGGATAAGT GCGGATAAGT TCTGTCTTAC CCGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG  
 3361 CCGTGGGCT GAACGGGGG TGAGCTATGA GAAAGCGCCA CGCTTCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAA  
 3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAA  
 3501 CCGGAGGCTC GGAACAGGAG AGCGCACGAG GAGCTTCCA GAGGAAACG GAGGAAACG GAGGAAACG GAGGAAACG  
 3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCTG GATGCTCTG GATGCTCTG GATGCTCTG  
 3641 AAAACGCCAG CAACGCGGC TTTTATACGTT TCTTGGGCTT TCTTGGGCTT TCTTGGGCTT TCTTGGGCTT TCTTGGGCTT  
 3711 TCGGTTATCC CCTGATTTCT TGGATTAACG TATTACGCG TATTACGCG TATTACGCG TATTACGCG TATTACGCG  
 3781 GGAACGACCG AGCGCAGCG GTACGAGAG GTACGAGAG GTACGAGAG GTACGAGAG GTACGAGAG GTACGAGAG  
 3851 CCGCGCGTTG GCCGATTCT TAAATGCAGT GGCACGACG GTTTCGCCG TGGAAAGCGG GCAGTGAGCG

FIG. 39D

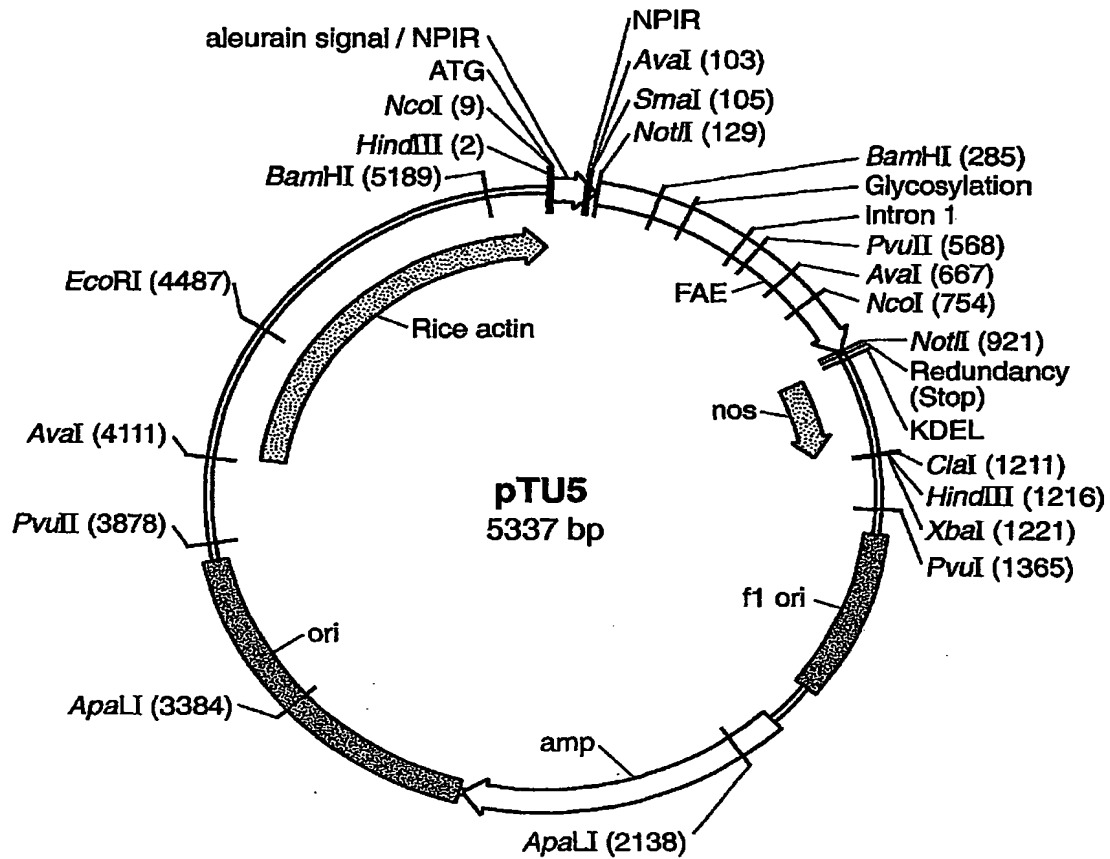


```

3921 CAACGGCAATT AATGTGAGTAT AGTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG TTTTCACACAG TTTTCACAGCTA TGACCATGAT TACGCCAAGC
                               KpnI
                               XhoI
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTGAGGTCA TTCTATATGCT
4131 TGAGAAGAGA GTCGGATAG TCCAAAATAA AACAAAGTA AGATTACCTG GTCAAAAGTG AAAACATCAG
4201 TTAAAAGGTG GTATAAGTAA AATATCGGTA ATAAAAGGTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAAAT TGAGGATGTT TTGTCTGGTAC TTTGTATACGT CATTTTGTG TGAATTGGTT TTTTAAGTTTA
4341 TTCGCGAATT GGAATGCAT ATCTGTATTT GAGTCGGTGT TTAAGTTCGT TGCTTTTGT AATACAGAGG
4411 GATTGTGATA AGAATATCT TTAAAAAACC CATATGCTAA TTGACATAA TTTTGTAGAA AATATATAT
                               EcoRI
4481 TCAGGCGAAT TCCCAATGA ACAATAATA GATTAAAAA GCTTGCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAGATA AACTTAGACT CAAACATTT ACAAAAACAA CCCCTAAAGT CCTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC AATCCACCCC AGTGCAGCCA
4691 ACTGGCAAT AGTCTCACC CCCGGCACTA TCACCGTGAG TTGTCCGCAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAAA AAGAAAGAAA AAAAAAGAAA AGAAAACAG CAGGTGGGTG GGGCCGGAAG GGGCCGGAAG
4831 AGCAGGAGG ATCGGAGCA GCGACGAGC CCGGCCCTCC CTCGGCTTC AAAGAAACGC CCCCATCGC
4901 CACTATATAC ATACCCCCC CTCTCCTCCC ATCCCCCCTC CCTTACCACC ACCACACCA CCACCTCTC
4971 CCCCCTCGCT GCCGACGAC GAGCTCCTCC CCCCCTCCCC TCCGCCGCC CCGGTAAACA CCCCCTCTC
5041 CTCTCTCTTC TTTCCTCGTT TTTTCTTTCG TCTCGGTCAC GATCTTTGGC CTGTGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCG GAGAGGGGCG GGATCTCGC GCTTGGCTCT CCGGGCGTGA
                               BamHI
                               BglI
5181 GTCGGCCCCG ATCTCGCGG GGAATGGGC TCCTCGGATGT AGATCTTCTT TCCTTCTTCT TTTTGTGTTA
5251 GAATTTGAAT CCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTTCATGAT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTGTAGC

```

FIG. 39E

**FIG.\_40A**

**HindIII NcoI**

~~~~~

1 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT
TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA

51 GGCCACGGCC GCCGTCGCCG TCGCCTCCTC CTCCTCCTTC GCCGACTCCA
CCGGTGCCGG CCGCAGCGGC AGCGGAGGAG GAGGAGGAAG CCGCTGAGGT

SmaI

~~~~~

**AvaI**

~~~~~

NotI

~~~~~

101 ACCCGGGCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC  
TGGGCCCCGGC CGGGCAGTGG CTGGCGCGCC GCGGAGGTG CGTCCCGTAG

151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC  
AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCG

201 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA  
ACGGATGCGG CTGGACACGT TGTAAAGCAG CTGATAATAG TTCCCTCTCT

**BamHI**

~~~~~

251 AAATTTACAA TTCTCAAAC T GACATTAACG GATGGATCCT CCGCGACGAC
TTTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GCGGCTGCTG

301 AGCAGCAAAG AAATAATCAC CGTCCTCCGT GGCACCTGGA GTGATACGAA
TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAAC TC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC
AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG

401 AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC
TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG

451 GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA GCCAGTATCC
CAGGTCTTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCATAGG

501 GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG
CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCGCGG AGGGACCGCC

PvuII

~~~~~

551 CACTCACTGC CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC  
GTGAGTGACG GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGGACATG

601 ACCTTCGGCG AACC GCGCAG CGGCAATCAG GCCTTCGCGT CGTACATGAA  
TGGAAGCCGC TTGGCGCGTC GCCGTTAGTC CGGAAGCGCA GCATGTACTT

**AvaI**

~~~~~

651 CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCGAGTAT TTCCGGGTCA
GCTACGGAAG GTTCGGAGCT CGGGTCCTATG CTGCGTCATA AAGGCCAGT

FIG. 40B

SUBSTITUTE SHEET (RULE 26)

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCAGGTGGA GCAGGGGTAC
 GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG

 NcoI
 ~~~~~  
 751 GCCCATGGCG GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCCAGAA  
 CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CGCGGGTCTT  
  
 801 CACATTTGTC TGCCTGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG  
 GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC CGGGTCCCCG  
  
 851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC  
 CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCTA CTGCTCGCCG  
  
 NotI  
 ~~~~~  
 901 GCATGCACCT GGCCGGTTCG GCGCGCGGAA CCACTGAAGG ATGAGCTGTA
 CGTACGTGGA CCGGCCAGCG CCGGCGCCTT GGTGACTTCC TACTCGACAT

 951 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT
 TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA

 1001 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT
 GGACAACGGC CAGAACGCTA CTAATAGTAT ATTAAAGACA ACTTAATGCA

 1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG
 ATTCGTACAT TATTAATTGT ACATFACGTA CTGCAATAAA TACTCTACCC

 1101 TTTTATGAT TAGAGTCCCG CAATFATACA TTTAATACGC GATAGAAAAC
 AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG

 1151 AAAATATAGC GCGCAAAC TA GGATAAATTA TCGCGCGCGG TGTATCTAT
 TTTTATATCG CGCGTTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

 XbaI
 ~~~~~  
 ClaI HindIII  
 ~~~~~  
 1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CCGTGGAGCT CCAATTCGCC
 CAATGATCTA GCTATTCGAA GATCTCGCCG GCCACCTCGA GGTAAAGCGG

 1251 CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC
 GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG

 1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
 CACTGACCC TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

 PvuII
 ~~~~~  
 1351 CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC  
 GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG  
  
 1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG  
 AAGGGTTGTC AACGCGTCGG ACTFACCGCT TACCCTGCGC GGGACATCGC

**FIG.\_40C**

SUBSTITUTE SHEET (RULE 26)

1451 GCGCATTAAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA  
CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA CTGGCGATGT

1501 CTTGCCAGCG CCCTAGCGCC CGCTCCCTTC GCTTTCCTCC CTTCTTTCT  
GAACGGTCCG GGGATCGCGG GCGAGGAAAG CGAAAGAAGG GAAGGAAAGA

1551 CGCCACGTTT GCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT  
GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG AGATTTAGCC CCCGAGGGAA

1601 TAGGGTTCCT ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGAT  
ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGGTT TTTTGAACFA

1651 TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTTCG  
ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAAGC

1701 CCCTTTGACG TTGGAGTCCA CGTCTTTTAA TAGTGGACTC TTGTTCCAAA  
GGGAAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG AACAAAGGTTT

1751 CTGGAACAAC ACTCAACCTT ATCTCGGTCT ATTCTTTTGA TTTATAAGGG  
GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAACT AAATATTCCC

1801 ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA  
TAAACCGGT AAAGCCGGAT AACCAATTTT TTAATCGACT AAATTGTTTTT

1851 ATTTAACGCG AATTTTAACA AAATATTAAC GCTTACAATT TAGGTGGCAC  
TAAATTGCGC TTAATAATTG TTTATAATTG CGAATGTTAA ATCCACCGTG

1901 TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC  
AAAAGCCCTT TTACACGCGC CTGCGGGATA AACAAATAAA AAGATTTATG

1951 ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCTTGATA AATGCTTCAA  
TAAGTTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAGTT

2001 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTGCGCCT  
ATTATAACTT TTTCTTCTC ATACTCATAA GTTGTAAGG CACAGCGGGA

2051 TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTGTCT CACCCAGAAA  
ATAAGGGAAA AAACGCCGTA AAACGGAAGG ACAAACCGA GTGGGTCTTT

ApaLI  
~~~~~

2101 CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT
GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG TGCTCACCCA

2151 TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC
ATGTAGCTTG ACCTAGAGTT GTCGCCATTC TAGGAACCTT CAAAAGCGGG

2201 CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG
GCTTCTTGCA AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC

2251 CGGTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA
GCCATAATAG GGCATAACTG CGGCCCGTTC TCGTTGAGCC AGCGGCGTAT

2301 CACTATTCTC AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA
GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTCGT

FIG. 40D

SUBSTITUTE SHEET (RULE 26)

2351 TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA
AGAATGCCTA CCGTACTGTC ATTCCTCTTAA TACGTCACGA CGGTATTGGT

2401 TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
ACTCACTATT GTGACGCCGG TTGAATGAAG ACTGTTGCTA GCCTCCTGGC

2451 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT
TTCCTCGATT GGCMAAAAAA CGTGTTGTAC CCCCTAGTAC ATTGAGCGGA

2501 TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG
ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC

2551 ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAAC
TGTGGTGCTA CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA

2601 GCGAACTAC TTA CTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA
CCGCTTGATG AATGAGATCG AAGGGCCGTT GTTAATTATC TGACCTACCT

2651 GCGGATATAA GTTGACAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT
CCGCTTATTT CAACGTCCTG GTGAAGACGC GAGCCGGGAA GGCCGACCGA

2701 GGTTTATTCG TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC
CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG

2751 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA
TAACGTCGTG ACCCCGGTCT ACCATTGCGG AGGGCATAGC ATCAATAGAT

2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG
GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCTTTATCT GTCTAGCGAC

2851 AGATAGGTGC CTCACTGATT AAGCATTTGGT AACTGTCAGA CCAAGTTTAC
TCTATCCACG GAGTGACTAA TTCGTAACCA TTGACAGTCT GGTTCAAATG

2901 TCATATATAC TTTAGATTGA TTTAAACTT CATTTTAAAT TTAAAGGAT
AGTATATATG AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCTTA

2951 CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG
GATCCACTTC TAGGAAAAAC TATTAGAGTA CTGGTTTTAG GGAATTGCAC

3001 AGTTTTTCGT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT
TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA

3051 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA
AGAACTCTAG GAAAAAAGA CGCGCATTAG ACGACGAACG TTTGTTTTTT

3101 ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC
TGGTGGCGAT GGTGCGCCACC AAACAAACGG CCTAGTTCTC GATGGTTGAG

3151 TTTTTCGAA GGTAACGGC TTCAGCAGAG CGCAGATACC AAATACTGTC
AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG TTTATGACAG

3201 CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC
GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCCTGA GACATCGTGG

3251 GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG
CGGATGTATG GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC

FIG. 40E

SUBSTITUTE SHEET (RULE 26)

3301 GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT
CGCTATTTCAG CACAGAATGG CCCAACCTGA GTTCTGCTAT CAATGGCCTA

ApalI

~~~~~

3351 AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT  
TTCCGCGTCG CCAGCCCAGC TTGCCCCCA AGCACGTGTG TCGGGTCGAA

3401 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG  
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTCGCA CTCGATACTC

3451 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC  
TTTCGCGGTG CGAAGGGCTT CCCTCTTTCC GCCTGTCCAT AGGCCATTCG

3501 GGCAGGGTCG GAACAGGAGA GCGCAGGAGG GAGCTTCCAG GGGGAAACGC  
CCGTCCCAGC CTTGTCTCTT CCGTGCTCC CTCGAAGGTC CCCCCTTTGCG

3551 CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC  
GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTCGCAG

3601 GATTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC  
CTAAAAACAC TACGAGCAGT CCCCCGCCT CGGATACCTT TTTGCGGTGCG

3651 AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT  
TTGCGCCGGA AAAATGCCAA GGACCGGAA ACGACCGGAA AACGAGTGT

3701 GTTCTTTTCT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT  
CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

3751 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG  
AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTGCTGGCT CGCGTCGCTC

3801 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC  
AGTCACTCGC TCCCTCGCCT TCTCGCGGGT TATGCGTTTG GCGGAGAGGG

PvuII

~~~~~

3851 CGCGCGTTGG CCGATTTCATT AATGCAGCTG GCACGACAGG TTTCCCGACT
GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGCTGTCC AAAGGGCTGA

3901 GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT
CCTTTCGCCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA

3951 TAGGCACCCC AGGCTTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG
ATCCGTGGGG TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC

4001 AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT
TTAACACTCG CCTATTGTTA AAGTGTGTCC TTTGTCGATA CTGGTACTAA

4051 ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC
TGCGGTTCGC GCGTTAATTG GGAGTGATTT CCCTTGTTTT CGACCCATGG

AvaI

~~~~~

4101 GGGCCCCCCC TCGAGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT  
CCCGGGGGGG AGCTCCAGTA AGTATACGAA CTCTTCTCTC AGCCCTATCA

**FIG. 40F**

SUBSTITUTE SHEET (RULE 26)

```

4151  CCAAAATAAA  ACAAAGGTAA  GATTACCTGG  TCAAAAGTGA  AAACATCAGT
      GGTTTTATTT  TGTTTCCATT  CTAATGGACC  AGTTTTCACT  TTTGTAGTCA

4201  TAAAAGGTGG  TATAAGTAAA  ATATCGGTAA  TAAAAGGTGG  CCCAAAGTGA
      ATTTTCCACC  ATATTCATTT  TATAGCCATT  ATTTTCCACC  GGGTTTCACT

4251  AATTTACTCT  TTTCTACTAT  TATAAAAATT  GAGGATGTTT  TGTCGGTACT
      TTAAATGAGA  AAAGATGATA  ATATTTTTAA  CTCCTACAAA  ACAGCCATGA

4301  TTGATACGTC  ATTTTGTAT  GAATTGGTTT  TTAAGTTTAT  TCGCGATTTG
      AACTATGCAG  TAAAAACATA  CTTAACCAAA  AATTCAAATA  AGCGCTAAAC

4351  GAAATGCATA  TCTGTATTTG  AGTCGGTTTT  TAAGTTCGTT  GCTTTTGTAA
      CTTTACGTAT  AGACATAAAC  TCAGCCAAAA  ATTCAAGCAA  CGAAAACATT

4401  ATACAGAGGG  ATTTGTATAA  GAAATATCTT  TAAAAAACCC  ATATGCTAAT
      TATGTCCTCC  TAAACATATT  CTTTATAGAA  ATTTTTTGGG  TATACGATTA

                                     EcoRI
                                     ~~~~~~

4451 TTGACATAAT TTTTGAGAAA AATATATATT CAGGCGAATT CCACAATGAA
 AACTGTATTA AAAACTCTTT TTATATATAA GTCCGCTTAA GGTGTTACTT

4501 CAATAATAAG ATTAAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTT
 GTTATTATTC TAATTTTATC GAACGGGGGC AACGTCGCTA CCCATAAAAA

4551 TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAAACATTTA CAAAAACAAC
 AGATCATTTT ATTTTCTATT TGAATCTGAG TTTTGTAAAT GTTTTGTGTTG

4601 CCCTAAAGTC CTAAGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA
 GGGATFTTCAG GATTTCTGGG TTCACGATAC GTGCTAGGTA TCGTTCGGGT

4651 GCCCAACCCA ACCCAACCCA ACCCACCCA GTGCAGCCAA CTGGCAAATA
 CGGGTTGGGT TGGGTGGGT TGGGTGGGT CACGTCGGT GACCGTTTAT

4701 GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC
 CAGAGGTGGG GGCCGTGATA GTGGCACTCA ACAGGCGTGG TGGCGTGCAG

4751 TCGCAGCCAA AAAAAAAAAA AGAAAGAAAA AAAAGAAAAA GAAAAACAGC
 AGCGTCGGTT TTTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTGTGTCG

4801 AGGTGGGTCC GGGTCGTGGG GGCCGGAAAA GCGAGGAGGA TCGCGAGCAG
 TCCACCCAGG CCCAGCACCC CCGGCCTTTT CGCTCCTCCT AGCGCTCGTC

4851 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
 GCTGCTCCGG GCCGGGAGGG AGGCGAAGGT TTCCTTTCGG GGGGTAGCGG

4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCAAC CCTACCACCA
 TGATATATGT ATGGGGGGGG AGAGGAGGGT AGGGGGGTTG GGATGGTGGT

4951 CCACCACCAC CACCTCCTCC CCCCTCGCTG CCGGACGACG AGCTCCTCCC
 GGTGGTGGTG GTGGAGGAGG GGGGAGCGAC GGCTGCTGC TCGAGGAGGG

5001 CCTCCCCCT CCGCGCGCGC CGGTAACCAC CCCGCCCTC TCCTCTTCT
 GGGAGGGGGA GGCAGCGCGC GCCATTGGTG GGGCGGGGAG AGGAGAAAGA

```

**FIG. 40G**

SUBSTITUTE SHEET (RULE 26)



5051 TTCTCCGTTT TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT  
AAGAGGCAAA AAAAAAAGCA GAGCCAGAGC TAGAAACCGG AACCATCAAA

5101 GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG  
CCCACCCGCT CTCGCCGAAG CAGCGGGTCT AGCCACGCGC CCTCCCCGCC

BamHI  
~~~~~

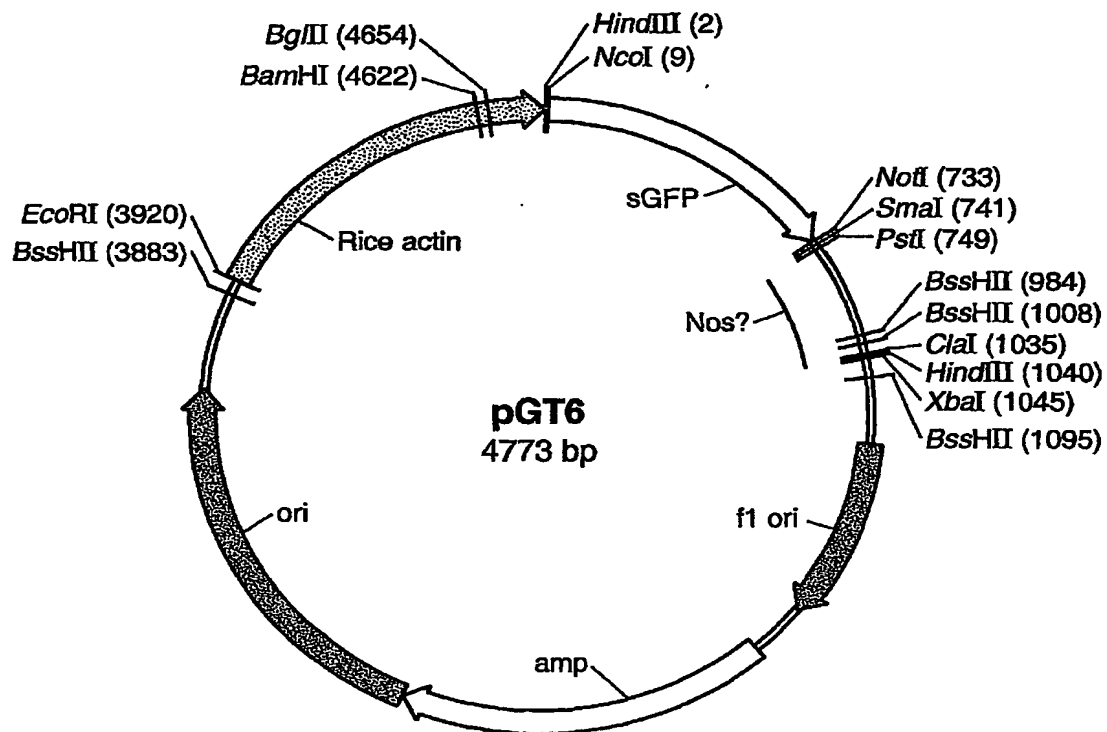
5151 GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG  
CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCCT AGGAGCGCCC

5201 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG  
CTTACCCCGA GAGCCTACAT CTAGAAGAAA GAAAGAAGAA AAACACCATC

5251 AATTTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT  
TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA AAGTACTAAA

5301 GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC  
CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCG

## FIG.\_40H

**FIG. 41A**

HindIII NcoI  
 ~~~~~  
 1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGTGTGTGCC CATCTGGTC GAGCTGGACG  
 TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACACAGG GTAGGACCAG CTCGACCTGC  
 71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGCGGAT GCCACCTACG GCAAGCTGAC  
 CGCTGCACCT GCGGTGTTC AAGTCGCACA GGGCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTGCACTG  
 141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCACCCC TCGTGACCAC CTTCACCTAC  
 GGACTTCAAG TAGACGTGGT GGCCGTTTGA CCGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG  
 211 GGGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGCACTT CTTCAGTCC GCCATGCCCG  
 CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGTGAA GAAATTGAGG CGGTACGGGC  
 281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CCGCAACTAC AAGACCCGCG CCGAGGTGAA  
 TTCCGATGCA GGTCTTCGG TGGTAGAAGA AGTTCTTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT  
 351 GTTCGAGGGC GACACCTTGG TGAACCGCAT CGAGTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC  
 CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCTCTCT GCCGTTGTAG  
 421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCTGACAAG CAGAAGAACG  
 GACCCCGTGT TCGACCTCAT GTTGATGTGG TCGGTGTGTC AGATATAGTA CCGGCTGTTC GTCTTCTTGC  
 491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTGCGCG ACCACTACCA  
 CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT  
 561 GCAGAACACC CCCATCGCGC ACGGCCCGGT GCTGCTGCCC GACACCACT ACCTGAGCAC CCAGTCCGCC  
 CGTCTTGTGG GGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGTTGA TGGACTCGTG GGTGAGGCGG  
 631 CTGAGCAAAG ACCCCACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA  
 GACTCGTTTC TGGGGTTGCT CTTGCGCGTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG. 41B

```

SmaI
~~~~~
NotI      PstI
~~~~~
701 CTCACGGCAT GGACGAGCTG TACAAGTAA CGGCGCGCCC GGGCTGCAGG GAAACCAC TG AAGGATGAGC
 GAGTGCCTGA CCTGCTCGAC ATGTTCAATT CGCGCGCGGG CCGACGCTCC CTTTGGTGAC TTCTCTACTCG

771 TGTAAAGRAG CAGATCGTTC AAACATTTGG CAATAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG
 ACATTTCTTC GTCTAGCAAG TTTGTAAACC GTTATTTCAA AGAATTTCTAA CTTAGGACAA CGGCCAGAAC

841 CGATGATTAT CATATAATT CTGTTGAATT ACGTTAAGCA TGTAAATAAT AACATGTAAT GCATGACGTT
 GCTACTAATA GTATATTAAA GACAACTTAA TGCAATTCGT ACATATTAA TTGTACATTA CGTACTGCAA

911 ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCCGCAATTA TACATTTAAT ACGCGATAGA AAACAAAATA
 TAAATACTCT ACCAAAAAT ACTAATCTCA GGGCGTTAAT ATGTAAATTA TCGCGTATCT TTTGTTTTAT

 ClaI HindIII
                                ~~~~~
                                XbaI
                                ~~~~~
 BssHII
                                ~~~~~
981 TAGCGCGCAA ACTAGGATAA ATTATCGCG CGGTGTGAT CTATGTTACT AGATCGATA GCTTCTAGAG
    ATCGCGCGTT TGATCCTATT TAATAGCGCG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC

                                BssHII
                                ~~~~~
1051 CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTACAA
 GCCGGCCACC TCGAGGTTAA GCGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCATAATGTT

1121 CGTCGTGACT GGGAAAACCC TGCGGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT
 GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG TGTAGGGGGA AAGCGGTCCA

1191 GCGGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACACTTGC GCCTTGAATG GCGAATGGGA
 CCGCATTATC GCTTCTCCGG GCGTGGCTAG CCGGAAGGGT TGTCAACGCG TCGGACTTAC CGCTTACCCCT

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC
 GCGCGGGACA TCGCCCGTA ATTGCGCGCG CCCACACCAC CAATGCGCGT CGCACTGGCG ATGTGAACGG

```

FIG. 41C

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCTCCCGTC  
TCGCGGGATC GCGGGCGAGG AAAGCGAAG AAGGCGGTG AAGAGCGGTG CAAGCGGCGG AAAGGGGCGAG  
1401 AAGCTCTAAA TCGGGGCTC CCTTTAGGGT TCCGATTTAG TCGTTTACGG CACCTCGACC CCAAAAAACT  
TTCGAGATTT AGCCCCGAG GGAATCCCA AGCTAAATC ACGAATGCC GTGGAGCTGG GGTTTTTTGA  
1471 TGATTAGGGT GATGTTTAC GTAGTGGCC ATCGCCCTGA TAGACGGTTT TTGCGCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGACT ATCGCAAA AAGCGGAAA CTGCAACCTC  
1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCTATCTCG GTCTATTTCTT  
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGAA  
1611 TTGATTATA AGGATTTTG CCGATTTTGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTAA  
AACTAAATAT TCCCTAAAAC GGCTAAAGCC GGATAACCAA TTTTTHACTC GACTAAATTG TTTTTHAAAT  
1681 CGCGAATTT AACAAATAT TAACGCTTAC AATTTAGGTG GCCTTTTCG GGGAAATGNG CGCGGAACCC  
GCGCTTAAA TTGTTTATA ATTCGGAATG TTAATCCAC CGTGAAGAAG CCCTTTTACAC GCGCCTTGGG  
1751 CTATTGTTT ATTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCTT GATAAATGCT  
GATAAACAAA TAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTTACGA  
1821 TCAATAATAT TGAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTTGTCG CCCTTATTCC CTTTTTTGCG  
AGTTATTATA ACTTTTTCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGAATAAGG GAAAAACGC  
1891 GCATTTGCG TTCTGTTT TTCTCACCCA GAAACGCTGG TGAAGTAAA AGATGCTGAA GATCAGTTGG  
CGTAAAACGG AAGGACAAA ACGAGTGGT CTTTGGGACC ACTTTCAATTT TCTACGACTT CTAGTCAACC  
1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTTC GCCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTCGCC ATTCTAGGAA CTCTCAAAAG CCGGGCTTCT  
2031 ACGTTTCCA ATGATGACA CTTTAAAGT TCTGCTATGT GCGCGGTAT TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAATTTCA AGACGATACA CCGCGCCATA ATAGGGCATA ACTGCGGCC  
2101 CAAGAGCAAC TCGGTGCGCG CATACACTAT TCTCAGATG ACTTGGTTGA GTACTCACA GTACAGAAA  
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

**FIG. 41D**

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC  
 TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG  
  
 2241 GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT  
 CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA AAAACGTGTT GTACCCCTTA  
  
 2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA  
 GTACATTGAG CGGAACCTAGC AACCTTGGC CTCGACTTAC TTCCGTAATG TTTGCTGCTC GCACGTGTGT  
  
 2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCC  
 GTACCGGACA TCGTTACCGT TGTTCGAACG CGTTTGATAA TTGACCCGCTT GATGAATGAG ATCGAAGGGC  
  
 2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TCGGCTCGGC CCTTCCGGCT  
 CGTTGTTAAT TATCTGACCT ACCTCCGCTT ATTTCAACGT CCTGCTGAG ACCTGAGCCG GGAAGGCCGA  
  
 2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG TATCATTTGCA GCACCTGGGC  
 CCGACCAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG  
  
 2591 CAGATGGTAA GCCCTCCCGT ATCTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA  
 GTCTACCAAT CGGGAGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCITT  
  
 2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACGT CAGACCAAGT TTACTCATAT  
 ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA  
  
 2731 ATACTTTAGA TTGATTTAAA ACTTCATTCT TAATTTAAAA GGATCTAGGT GAAGATCCTT TTTGATAATC  
 TATGAATCT AACTAAATTT TGAAGTAAAA ATTAAATTTT CCTAGATCCA CTTCTAGGAA AAACATATTAG  
  
 2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CATTCCACTG AGCTCAGAC CCCGTAGAAA AGATCAAAGG  
 AGTACTGGTT TTAGGGAATT GCACCTAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC  
  
 2871 ATCTTCTTGA GATCCTTTTT TTCTGGCGGT AATCTGCTGC TTGCAAAACA AAAAACACC GCTACCAGCG  
 TAGAAGAAT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGT TTTTGTGGTG CGATGGTCCG  
  
 2941 GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGTAAAC TGGCTTCAGC AGAGCGCAGA  
 CACCAAAACA ACGGCCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTT ACCGAAAGTCG TCTCGCGTCT

**FIG. 41E**

3011 TACCAAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTTCAAG AACTCTGTAG CACCGCCTAC  
 ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG  
 3081 ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG  
 TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGCCCCAAC  
 3151 GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTGCG GGTTCGTGC ACACAGCCCA  
 CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCAGG TGTGTGGGT  
 3221 GCTTGGAGCG AACGACCTAC ACCGAACCTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  
 CGAACCTCGC TTGCTGGATG TGGCTTGAAT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG  
 3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGAGCTT  
 GCTTCCCTCT TTCCGCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTG CTCTCGCGTG CTCCTCGAA  
 3361 CCAGGGGGA AGCCTTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT  
 GGTCCCCCTT TCGGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAATC GCAGCTAAAA  
 3431 TGTGATGCTC GTCAGGGGG GGGAGCCTAT GGAATAACGC CAGCAACGCG GCCTTTTATC GGTTCCTGGC  
 AACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTGGG GTCGTTGCG CGGAAAAATG CCNAGGACCG  
 3501 CTTTGTCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGAT CTGTGGATAA CCGTATTACC  
 GAAAACGACC GGAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATAATGG  
 3571 GCCTTTGAGT GAGTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG  
 CGGAAACTCA CTCGACTATG GCGAGCGGC TCGGCTTGCT GGTCTCGCTC GCTCAGTCAC TCGCTCCTTC  
 3641 CGGAAGAGCG CCCAATACGC AAACCGCTC TCCCCGCGG TTGGCCGATT CATTAATGCA GCTGGCACGA  
 GCCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGCGCGC AACCGGCTAA GTAAATTACGT CGACCGTGCT  
 3711 CAGGTTTCCC GACTGGAAAG CCGGCAGTGA GCGCAACGCA ATTAATGTA GTTAGCTCAC TCATTAGGCA  
 GTCCAAAGG CTGACCTTTC GCCGCTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT  
 3781 CCCCAGGCTT TACACTTAT GCTTCCGGCT CGTATGTTGT GTGAAATTGT GAGCGGATAA CAATTTCACA  
 GGGTCCGAA ATGTGAATA CGAAGCCGA GCATACAACA CACCTTAACA CTCGCCCTATT GTTAAAGTGT

**FIG. 41F**

```

 BssHII EcoRI
      ~~~~~
3851 CAGGAACACAG CTATGACCAT GATTACGCCA AGCGGCAAT TAACCTCAC TAAAGGAAC AAAAGCTGGA
      GTCCTTTGTC GATACTGGTA CTAATGCGGT TCGCGCGTTA ATTGGGAGTG ATTTCCCTTG TTTTCGACCT

      EcoRI
      ~~~~~
3921 ATTCCACAAT GAACAATAAT AAGATTAAAA TAGCTTGCCC CCGTTGCAGC GATGGGTATT TTTTCTAGTA
 TAAGGTGTTA CTTGTTATTA TTCTAATTTT ATCGAACGGG GGCAACGTCG CTACCCATAA AAAAGATCAT

3991 AAATAAAGA TAAACTTAGA CTCAAAACAT TTACAAAAC AACCCCTAAA GTCTAAAGC CCAAAGTGT
 TTTATTTTCT ATTTGAATCT GAGTTTGTGA AATGTTTGTG TTGGGGATT TTTTCACGA

4061 ATGCACGATC CATAGCAAGC CCAAGCCAAAC CCAACCCACC CCAAGTCAGC CCAACTGGCAA
 TACGTGCTAG GTATCGTTTCG GGTGCGGTG GGTGGGTG GGTACGTCG GTTGACCGTT

4131 ATAGTCTCCA CCCCAGGCAC TATCACCGTG AGTTGTCCGC ACCACCGCAC GTCCTGCAGC CAAAAAAA
 TATCAGAGGT GGGGGCCGTG ATAGTGGCAC TCAACAGGCG TGGTGGCGTG CAGAGCGTCG GTTTTTTT

4201 AAAAGAAAGA AAAAAGAA AAAAAGAAAGC AGCAGGTGG TCCGGGTCTG GGGGGCCGGA AAAGCGAGGA
 TTTTCTTCT TTTTCTTCT TTTTCTTCT TCGTCCACCC AGGCCACGA CCCCAGCCT TTTTCGCTCT

4271 GGATCGCGAG CAGCGACGAG GCCCGGCCCT CCTCCGCTT CCAAGAAAC GCTCCCTCAT GCCACTATAT
 CCTAGCGCTC GTCGCTGCTC CCGGCCGGA GGGAGGCGAA GGTTCCTTTG CGGGGGTAG CGGTGATATA

4341 ACATACCCCC CCTCTCTCT CCAATCCCCC AACCTTACCA CCACCACTCC CACCCCTCTG TCCCCCTCG
 TGTATGGGG GGGAGAGGAG GGTAGGGGG TTGGGATGGT GGTGTTGTTG GTGTTGAGG AGGGGGAGC

4411 CTGCCGGACG ACGAGTCTCT CCCCCTCTCC CCTCCGCGC CGCCGGTAA CACCCCGCCC CTCTCTCTCT
 GACGGCTGC TGCTCGAGGA GGGGGGAGGG GAGGCGGCG GCGGCCATTG GTGGGGCGGG GAGAGGAGA

4481 TCTTCTCTCG TTTTCTTTT CGTCTCGGTC TCGATCTTTG GCCTTGGTAG TTGGGTGGG CGAGAGCGGC
 AGAAAGAGGC AAAAAAAA GCAGAGCCAG AGCTAGAAAC CGGAACCATC AAACCAACC GCTCTCGCCG

4551 TTGCTCGCCC AGATCGGTGC GCGGAGGGG CGGGATCTCG CGGTGGCGT CTCCGGGCCC GAGTCGGCCC
 AAGCAGCGGG TCTAGCCACG CGCCCTCCCC GCCCTAGAGC GCGGACCGCA GAGGCCCGCA CTCAGCCCGG

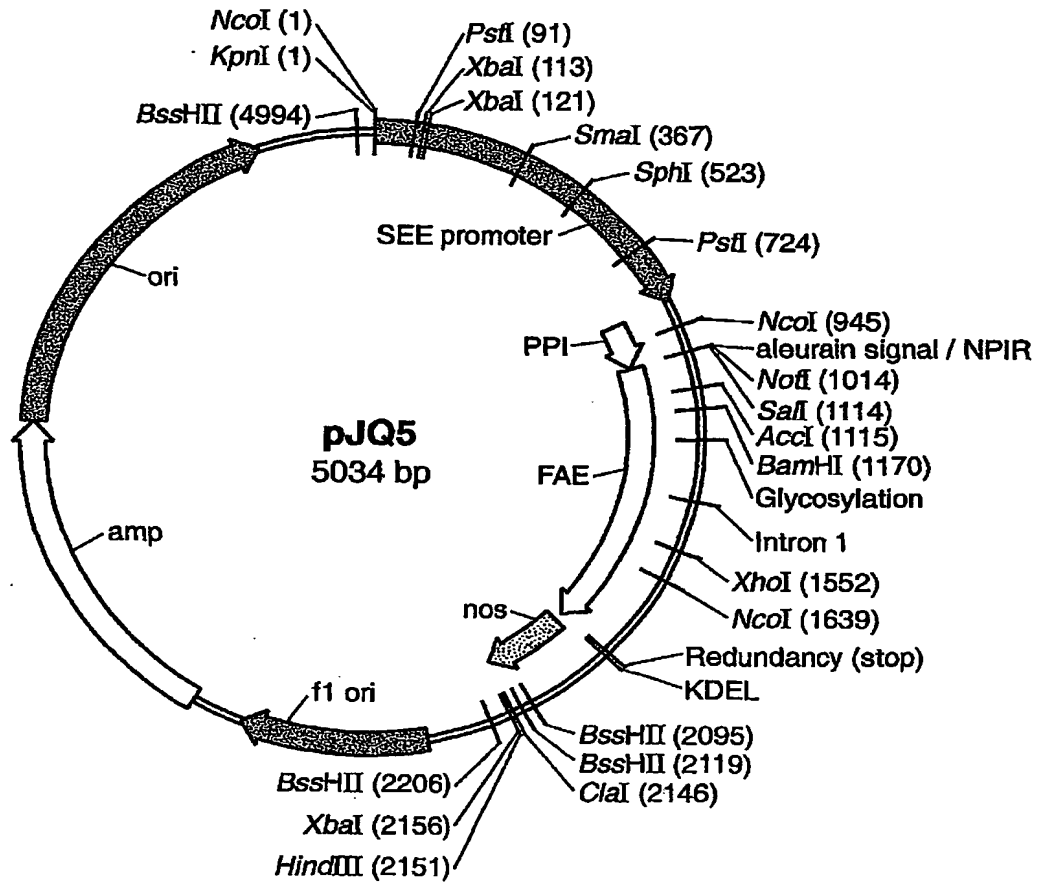
```

**FIG. 41G**



|      | BamHI                                                                                                                                                         | BglII |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|      | ~~~~~                                                                                                                                                         | ~~~~~ |
| 4621 | GGATCCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTTC TTTCTTTCTT CTTTTGTGG TAGAAATTGA<br>CCTAGGAGCG CCCCTTACCC CGAGAGCCTA CATCTAGAAG AAAGAAAGAA GAAAAACACC ATCTTAAACT   |       |
| 4691 | ATCCCTCAGC ATTGTTTCATC GGTAGTTTTT CTTTTCATGA TTTGTGACAA ATGCAGCCTC GTGCGGAGCT<br>TAGGGAGTCG TAACAAGTAG CCATCAAAA GAAAAGTACT AAACACTGTT TACGTCGGAG CACGCCCTCGA |       |
| 4761 | TTTTTGTAGG TAG<br>AAAAACATCC ATC                                                                                                                              |       |

**FIG.\_41H**

**FIG. 42A**

NcoI  
 ~~~~~  
 KpnI  
 ~~~~~  
 1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAATATA TCGAATATC ACCATTGGCT ACAATATCTG  
 GTACCCGGTC CATATTAAATA CCTATATAG TTCGTTTATT AGCTTTATAG TGCTAACCGA TGTATATAGAC  
  
 PstI  
 ~~~~~  
 XbaI  
 ~~~~~  
 XbaI  
 ~~~~~  
 71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC  
 TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAAACA TAGATCTTGA GATCTATCGT GTCGGTGTCTG  
  
 141 ACCTACAGGA GTCCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT  
 TGGATGTCTT CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCCTCG AGAAGGATG GAGGACTGCA  
  
 211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGGGCTC CCAACAAAAT ATCGTCCCCC  
 ACGGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGCGCGAG GGTGTGTTTA TAGCAGGGGG  
  
 281 ATGTCCTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTGCT GAATCTCGCT TCCACTGGCC  
 TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACACGA CTTAGAGCGA AGGTGACCGG  
  
 SmaI  
 ~~~~~  
 351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGBA  
 TTAGTTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCCT  
  
 421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT  
 ACCTTTTCTT GGAGGCAACG AGCCTACTCA GTCCGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA  
  
 SphI  
 ~~~~~  
 491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATGAC TTTATTTTTG  
 AGCTAAAGG GGTATCTCTT TTCAATATCGT ACGAAAGCCC AAAACAAACC GAATTAACAG AAATAAANAC  
  
 561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAF CGAGACGGAT AATAGGCTGG  
 AACCTCAACT TAGGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCCTA TTATCCGACC

FIG. 42B

```

631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTTACA GCTGGAGGCT
 GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

 PstI
          ~~~~~
701  TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
    AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCTGTTACC ACTTCGCACC GCTACTCTAC CCATATTTTG

771  CCCCAGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCCTGC TCCCCCTGCC GGACGACCCA
    GGGGCCGTGG CCTTGCCTC GAGGCGGAT GGTCAATGGA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841  GTAAATACT GTTGCCCACT CGCCGCGGAG ATGCMCGTC ACAAGGAGT SAAC TTCGTS GCCTACCTCC
    CATTTATGA CAACGGGTGA GCGGCCGCTC TACCKGCACG TGTTCCTCCA TTGAAGCAS CGGATGGAGG

          NcoI
          ~~~~~
911 TGATCGTCT CGGCCCTCCTC TTGCTCGTST CCGCCATGGA GCACGTGGAG GCCAAGGCTT GCACCCKCGA
 ACTAGCAGBA GCCGGAGGAG AACGAGCASA GCGCGTACCT CGTGCACCTG CGGTTCCGGA CGTGGGCGCT

 NotI
          ~~~~~
981  GTGCGGCAAC CTCGGCTTCG GCATCTGCCC GCGGCGCGCC TCCACGCGAG GCATCTCCGA AGACCTCTAC
    CACGCCGTTG GAGCCGAAGC CGTAGACGGG CCGCCGGCGG AGGTGCGTCC CGTAGAGGCT TCTGGAGATG

          SalI
          ~~~~~
 AscI
          ~~~~~
1051 AGCCGTTT AGCGAAATGGC CACTATCTCC CAAGCTGCCCT ACGCCGACCT GTGCAACATTT CCGTCGACTA
    TCGGCAAAATC AGCTTTACCG GTGATAGAGG GTTCGACGGA TCGCGGCTGA CACGTTGTAA GGCAGCTGAT

          BamHI
          ~~~~~
1121 TTATCAAGGG AGAGAAATTT TACAATCTTC AAAGTGACAT TAACGATGG ATCCTCCGG ACGACAGCAG
 AATAGTTCCC TCCTCTTTAA ATGTTAAGAG TTTGACTGTA ATTGCTTACC TAGGAGGCGC TGCTGTCTGC

```

FIG. 42C

```

1191 CAAAGAAATA ATCACCGTCT TCCGTGGCAC TGGTAGTGAT ACGAATCTAC AACTCGATAC TAACATACACC
 GTTCTTTTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TGCTTAGATG TTGAGCTATG ATTGATGTGG

1261 CTCACGCCCTT TCGACACCCCT ACCACAATGC AACGGTTGAG AAGTACACGG TGGATATTAT ATTGGATGGG
 GAGTGGCGAA AGCTGTGGGA TGGTGTACG TTGCCAACAC TTTCATGTGCC ACCATAAATA TAACCTACCC

1331 TCTCCGTCCA GGACCAAGTC GAGTCGCTTG TCAAAACAGCA GGTAGCCAG TATCCGGACT ACGCGCTGAC
 AGAGGCAGGT CCTGGTTTCAG CTCAGCGAAC AGTTTGTCTGT CCAATCGGTC ATAGGCCTGA TGCGCCGACTG

1401 CGTGACCGGC CACKCCCTCG GCGCCTCCCTT GCGGGCACTC ACTGCCGCCC AGCTGTCTGC GACATACGAC
 GCACGTGCCG GTGMMGGGAGC CCGCGAGGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTATGCTG

1471 AACATCCGCC TGTACACCTT CCGCGAACCG CGCAGCGGCA ATCAGGCCCTT CGCTCTGTAC ATGAACGATG
 TTGTAGGCGG ACATGTGGAA GCCGCTTGGC CGCTCGCCGT TAGTCCGGAA GCGCAGCATG TACTTGCTAC

 XhoI
      ~~~~~
1541 CCTTCCAAGC CTCGAGCCCA GATACGACGC AGTATTTCCG GGTCACTCAT GCCAACGACG GCATCCCAA
      GGAAGGTTCT GAGCTCGGGT CTATGCTGCG TCATAAAGGC CCAGTGAGTA CGGTGTCTGC CGTAGGGTTT

      NcoI
      ~~~~~
1611 CCTGCCCCCG GTGGAGCAGG GGTACGCCCA TGGCGGTGTA GAGTACTGGA GCGTTGATCC TTACAGCGCC
 GGACGGGGGC CACCTCGTCC CCATGCGGGT ACCGCCACAT CTCATGACCT CGCAACTAGG AATGTCCGG

1681 CAGAACACAT TTGTCCTGCAC TGGGGATGAA GTGCAGTCTT CACGTCACGA CACTCCGGGT CCGCCTGTC
 GTCTTGTTGA AACAGACGTG ACCCTACTT CACGTCACGA CACTCCGGGT CCGCCTGTC CCACACTTAT

1751 ATGCGCACAC GACTTATTTT GGGATGACGA GCGGAGCCCTG TACATGGTGA TCAGTCATTT CAGCCTCCCC
 TACGCGTGTG CTGAATATAA CCCTACTGCT CGCCTCGGAC ATGTACCACT AGTCAGTAAA GTCGGAGGGG

1821 GAGTGTACCA GGAAGATGG ATGTCCTGGA GAGGGGGCCG CGTAACCACT GAAGGATGAG CTGTAAAGAA
 CTCACATGGT CCTTCTTACC TACAGGACCT CTCCCCCGGC GCATTTGGTGA CTTCTTACTC GACATTTCTT

```

**FIG. 42D**

```

1891 GCAGATCGTT CAAACATTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TCCGGTCTT GCATGATTA
CGTCTAGCAA GTTTGTAAAC CGTTATTCA AAGAATTCTA ACTTAGGACA ACGCCAGAA CGCTACTAAT

1961 TCATATAAAT TCTGTGGAAT TACGTTAAGC ATGTAATAAT TAACATGTAA TGCATGACGT TATTATGAG
AGTATATATA AGACAACCTA ATGCAATTG TACATTATTA ATTGTACATT ACGTACTGCA ATAAATACTC

 BssHII
                ~~~~~

2031 ATGGGTTTTT ATGATTAGAG TCCCGCAATT ATACATTTAA TACGGGATAG AAAACAAAAT ATAGCGCGCA
TACCCAAAAA TACTAATCTC AGGCGGTAA TATGTAAATT ATGGCTATC TTTTGTTTTA TATCGCGCGT

                BssHII
                ~~~~~
 ClaI HindIII
                ~~~~~

2101 AACTAGGATA AATTATCGCG CGCGGTGTC TCTATGTTAC TAGATCGATA AGCTTCTAGA GCGGCCGGTG
TTGATCCCTAT TTAATAGCGC GCGCCACAGT AGATACAATG ATCTAGCTAT TCGAAGATCT GCGCGGCCAC

                BssHII
                ~~~~~

2171 GAGCTCCAAT TCGCCCTATA GTGAGTCGTA TTACGCGCGC TCACAGGCCG TCGTTTACA ACGTCGTGAC
CTCGAGGTTA AGCGGGATAT CACTCAGCAT AATGCGCGCG AGTGACCGGC AGCAAAATGT TGCAGCACTG

2241 TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG CACATCCCCC TTTCGCCAGC TGGCGTAATA
ACCTTTTGG GACCGCAATG GGTGGAATTA GCGGAACGTC GTGTAGGGG AAGCGGTG ACCGCATTAT

2311 GCGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGCG CAGCCTGAT GCGAATGGG ACGCGCCCTG
CGCTTCTCCG GCGGTGGCTA GCGGGAAGGG TTGTCAACGC GTCGGACTTA CCGCTTACCC TCGCGCGGAC

2381 TAGCGGCGCA TTAAGCGCGG CGGTGTGGT GGTACGCGC AGCGTGACC CTACACTTGC CAGCGCCCTA
ATCGCCGCGT AATTCGCGCC GCCCACACCA CCAATGCGCG TCGCACTGCG GATGTGAACG GTCGCGGGAT

2451 GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTCTTCGCCA CGTTCGCGCG CTTTCCCGGT CAAGCTCTAA
CGCGGCGGAG GAAAGCGAAA GAAGGAAGG AAGAGCGGT GCAAGCGGCC GAAAGGGGCA GTTCGAGATT

```

**FIG.-42E**

2521 ATCGGGGGCT CCCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAC TTGATTAGGG  
TAGCCCCCGA GGGAAATCCC AAGGTAAAT CACGAANTGC CGTGGAGCTG GGGTTTTTGG AACTAATCCC

2591 TGATGGTTCA CGTAGTGGG CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA GTCCACGTTT  
ACTACCAAGT GCATCACCCG GTAGCGGAC TATCTGCCAA AAGCGGGA ACTGCAACCT CAGGTGCAAG

2661 TTTAATAGTG GACTCTTGT CCAAACTGGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTAT  
AAATTATCAC CTGAGAACA GGTTTGACCT TGTGTGAGT TGGGATAGAG CCAGATAAGA AACTAAATA

2731 AAGGATTTT GCCGATTTCG GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTA ACGCGAATTT  
TTCCCTAAAA CGGCTAAAGC CGGATAACCA ATTTTTTACT CGACTAAAT GTTTTAAAT TGGCTTAA

2801 TAACAAAATA TTAACGCTTA CAATTTAGGT GGCACITTTT GGGGAAATGT GCGCGCTTGG CCTATTGT  
ATTGTTTAT ATTGCGAAT GTTAAATCCA CCGTGAAAG CCCCTTTACA CCGCGCTTGG GGTAAACAA

2871 TATTTTCTA AATACATTCA AATATGTATC CGTCTATGAG ACAATAACCC TGAATAATGC TTCAATAATA  
ATAAAAAGAT TTATGTAAAT TTATACATAG GCGAGTACTC TGTATTGGG ACTATTACG AAGTTATTAT

2941 TTGAAAAAGG AAGAGTATGA GTATTCAACA TTTCCTGTTC GCCCTTATTC CCTTTTTCG GGCATTTTCG  
AACTTTTTC TTCTCATACT CATAAGTTGT AAAGGCACAG CCGGAATAAG GGAATAAACG CCGTAAAAACG

3011 CTTCCGTGTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG  
GAAGGACAAA AACGAGTGGG TCTTTGCGAC CACTTTCATT TTCTACGACT TCTAGTCAAC CCACGTGCTC

3081 TGGGTTACAT CGAACTGGAT CTCAACAGCG GTAAAGATCCT TGAGAGTTT CGCCCCGAAG AACGTTTCC  
ACCCAATGTA GCTTGACCCTA GAGTTGTGCG CATCTAGGA ACTCTCAAAA CCGGGCTTC TTGCAAAAGG

3151 AATGATGAGC ACTTTTAAAG TTCTGCTATG TGGCGCGGTA TTATCCCGTA TTGACGCCGG GAAGAGCAA  
TTACTACTCG TGAAAAATTC AAGACGATAC ACCGCGCCAT AATAGGGCAT AACTGCGGCC CGTTCTCGTT

3221 CTCGGTCGCC GCATACACTA TTCTCAGAAAT GACTTGGTTG AGTACTCAC AGTCACAGAA AAGCATCTTA  
GAGCCAGCGG CGTATGTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTCGTAGNAT

3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCAATGAGT GATAACACTG CGGCCAACTT  
GCCTACCGTA CTGTCAATCT CTTAATACGT CACGACGGTA TTGGTACTCA CTATTGTGAC GCCGGTTGAA

**FIG. 42F**

3361 ACTTCTGACA ACGATCGGAG GACCGAAGGA GCTAAACCGCT TTTTTCACACA ACATGGGGGA TCATGTAAC T  
 TGAAGACTGT TGCTAGCCTC CTGGCTTCCT CGATTGGCGA AAAAACGTGT TGTACCCCTT AGTACATTGA  
 3431 CGCCTTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC ACGATGCCCTG  
 GCGGAAC TAG CAACCCCTGG CCTCGACTTA CTTTCGGTATG GTTTGCTGCT CCGACTGTGG TGCTACGGAC  
 3501 TAGCAATGGC AACAACTGT CGCAAACTAT TAACTGGCGA ACTACTTACT CTAGCTTCCC GGCAACAATT  
 ATCGTTACCG TTGTGTCAAC GCGTTTGATA ATTGACCGCT TGATGAATGA GATCGAAGGG CCGTTGTATA  
 3571 AATAGACTGG ATGGAGGCGG ATAAAGTTC ATAAAGTTC AGGACCACTT CTGCGCTCGG CCTTCCGGC TGGCTGGTT  
 TTATCTGACC TACCTCCGCC TATTTCAACG TCCTGGTGA GACGCGAGCC GGGAGGGCCG ACCGACCAAA  
 3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGTCTCGCG GTATCATTC AGCACTGGGG CCAGATGGTA  
 TAACGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGCGC CATAGTAACG TCGTGACCCC GGTCTACCAT  
 3711 AGCCCTCCCG TATCGTAGTT ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT  
 TCGGGAGGGC ATAGCATCAA TAGATGTGCT GCCCCTCAGT CCGTTGATAC CTACTTGCTT TATCTGTCTA  
 3781 CGCTGAGATA GGTGCCCTAC TGATTAAGCA TTGGTAACTG TCAGACCAAG TTTACTCATA TATACTTTAG  
 GCGACTCTAT CCACGGAGTG ACTAATTCGT AACCATTTGAC AGTCTGGTTC AAATGAGTAT ATATGAAATC  
 3851 ATTGATTTAA AACTTCAATTT TTAATTTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCATGACCA  
 TAACTAAATT TTGAAGTAAA AATTAATTT TCCTAGATCC ACTTCTAGGA AAAACTATTA GAGTACTGGT  
 3921 AAATCCCTTA ACGTGAGTTT TCGTCCACT GAGCGTCAGA CCCCCTAGAA AAGATCAAAG GATCTTCTTG  
 TTTAGGGAAT TGCACCTCAA AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAAGAAC  
 3991 AGATCCCTTT TTTCTGCGCG TAATCTGCTG CTTGCAAAACA AAAAACACAC CGCTACCAGC GGTGGTTTGT  
 TCTAGGAAAA AAAGACGCGC ATTAGACGAC GAACGTTTGT TTTTGTGGTG GCGATGGTTC CCACCAAAACA  
 4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAAATA  
 AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGTCTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT  
 4131 CTGTCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA GAACCTGTA GCACCGCCTA CATACCTCGC  
 GACAGGAAGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG

**FIG. 42G**



4201 TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA  
AGACGATTAG GACAATGGTC ACCGACGACG GTCACCGCTA TTCAGGCACAG AATGGCCCAA CCTGAGTTCT

4271 CGATAGTTAC CGGATAAGGC GCAGCGGTG GGTGAACGG GGGTTCTGT CACACAGCCC AGCTTGGAGC  
GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGGC CCCCAAGCAC GTGTGTGGG TCGAACCTCG

4341 GAACGACCTA CACCGAACTG AGATACTAC AGCTGAGCT ATGAGAAAAGC GCCACGCTTC CCGAAGGGAG  
CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTCG CGGTGCGAAG GGCTTCCCTC

4411 AAAGCGGAC AGGTATCCGG TAAGCGGAG GGTGGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGA  
TTTCCGCCCTG TCCATAGGCC ATTCCCGTC CCAGCTTGT CCTCTCGGT GCTCCCTCGA AGTCCCTCT

4481 AACGCTGGT ATCTTTATAG TCCTGTGCGG TTTCCGCCACC TCTGACTTGA GCCTCGATT TTTGTGATGCT  
TTGCGGACCA TAGAATATC AGGACAGCC AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA

4551 CGTCAGGGG GCGGAGCCTA TGGAAAAAG CCAGCAACGC GGCTTTTITA CGGTCTCTGG CTTTTTGTG  
GCAGTCCCC CGCTTCGGAT ACCTTTTTC GGTGTTTGC CCGGAAAAAT GCCAAGGACC GGAAAAACGAC

4621 GCCTTTTGT CACATGTTCT TTCCTGGGT ATCCCTGAT TCTGTGATA ACCGTATTAC CGCTTTTGG  
CGGAAAAACGA GTGTACAAGA AAGGACGCA TAGGGACTA AGACACCTAT TGGCATATG GCGGAAACTC

4691 TGAGCTGATA CCGCTCGCG CAGCCGAACG ACCGAGCGCA GCGAGTCAAT GAGCGAGGAA GCGGAAGAGC  
ACTCGACTAT GCGGAGCGGC GTCGGCTTGC TGGCTCGCTG CGCTCAGTCA CTCGCTCTCTT CGCCTTCTCG

4761 GCCCAATACG CAACCCGCT CTCGCCGCG GTTGGCCGAT TCATTATGC AGCTGGCACG ACAGGTTTCC  
CGGTTATGC GTTTGGCGA GAGGGGCGG CAACCGGCTA AGTAATTACG TCGACCGTGC TGTCCAAAGG

4831 CGACTGGAA GCGGGCAGT AGCGCAACGC AATTAATGT AGTTAGCTCA CTCATTAGGC ACCCCAGGCT  
GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAC TCAATCGAGT GAGTAATCCG TGGGGTCCGA

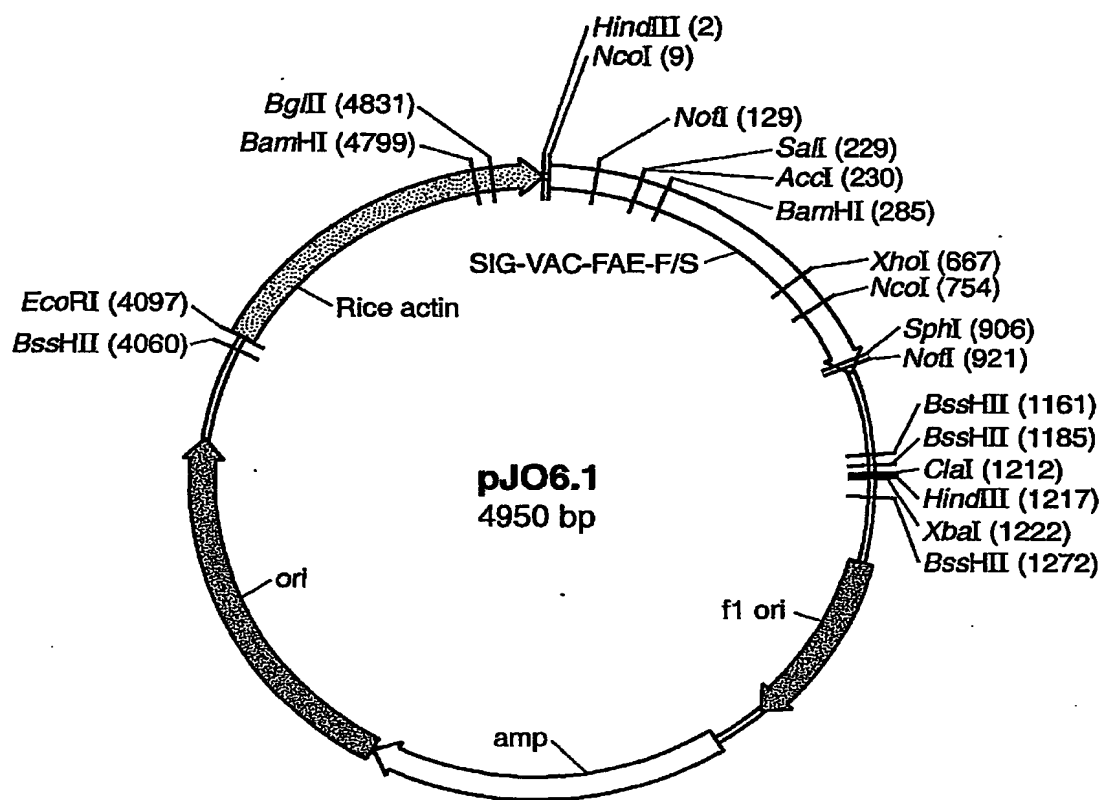
4901 TTACACTTTA TGCTTCCGG TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAATTTCAC ACAGGAAACA  
AATGTGAAT ACGAAGGCCG AGCATAAC ACACCTTAACT ACTCGCTTAT TGTAAAGTG TGTCTTTTGT

FIG.\_42H

4971

	BSSHII		NcoI
	~~~~~		KpnI
		~~~~~	
	GCTATGACCA	TGATTACGCC	AAGCGCGCAA
	CGATACTGGT	ACTAATGCCG	TTCGCGCGTT
			CTAAAGGGA
			CAAAAGCTGG
			GTAC
			GATTTCCGAC
			CATG

FIG.\_42I

**FIG. 43A**

HindIII NcoI  
 ~~~~~  
 1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTGCGC  
 TTGGAATGGT ACCGGGTGCG GCGCAGGAG GAGGACCGC AGCGGCACGA CCGGTGCCG CGGCAGCGC

NotI  
 ~~~~~  
 71 TCGCCTCCTC CTCTCCTTC GCCACTCCA ACCGATCCG GCCGTCACTC GACCGCGCGG CCGCCTCCAC  
 AGCGAGGAG GAGGAGGAG CCGCTGAGT TGGGCTAGG CCGGCAGTGG CCGCGCGCGC GCGGAGGTG

141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC  
 CGTCCCGTAG AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGTTG ACGGATGCGG

Sall  
 ~~~~~  
 AccI  
 ~~~~~  
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTACG  
 CTGGACACGT TGTAAAGCAG CTGATAATAG TTCCCTCTCT TTTAAATGTT AAGAGTTTGA CTGTAAATGC

BamHI  
 ~~~~~  
 281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACCTGGTA GTGATACGAA  
 CTACCTAGGA GCGCTGCTG TCGTCGTTTC TTTATTATAG GCAGAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAATC GATACTAACT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAGTA  
 AGATGTTGAG CTATGATGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG TTACGTTGCC AACACTTCAT

421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
 GTGCCACCTA TAATATAACC TACCAGAGG CAGGTCCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT

491 GCCAGTATCC GGACTACGG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCCTGCG  
 CCGTCAATAG CCTGATGCGC GACTGGCACT GCGCGGTGAG GAGCCGCGG AGGACCGCC GTGAGTGCAG

561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCTGTAC ACCTTCGGCG AACCGCGCAG CGGCATCAG  
 GCGGTCGAC AGACGCTGTA TGCTGTTGTA GCGCGACATG TGAAGCCGC TTGGCGCGTC GCCGTTAGTC

**FIG. 43B**

XhoI  
~~~~~

631 GCCTTCGCGT CGTACATGAA CGATGCCCTTC CAAAGCCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA  
CGGAAGCGCA GCATGTACTT GCTACGGAAG GTTCGGAGCT CGGCTCTATG CTGCGTCATA AAGGCCAGT

NcoI  
~~~~~

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCGCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG CCGGTACCGC CACATCTCAT

771 CTGGAGCGTT GATCCTTACA GCGCCCAGAA CACATTTGTC TGCACCTGGG ATGAAGTGCA GTGCTGTGAG  
GACCTCGCAA CTAGGAATGT CGCGGGTCTT GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC

SphI  
~~~~~

841 GCCCAGGGCG GACAGGGTGT GAATAATGC CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT  
CGGGTCCCGC CTGTCCACCA CTTATTACGC GTGTCTGAA TAAACCCCTA CTGCTCGCCG CGTACGTGGA

NotI  
~~~~~

911 GGCCGGTCCG GGCCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAAA CATTTGGCAA  
CCGGCCAGCG CCGGCGCCTT TGGTGACTTC CTACTCGACA TTCTTTCGTC TAGCAAGTTT GTAAACCGTT

981 TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCGA TGATTATCAT ATAAATTTCTG TTGAATTACG  
ATTTCAAAGA ATCTTAACCT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC

1051 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGGA TTAGAGTCCC  
AATTGCTACA TTATTAATTG TACATTACGT ACTGCAATAA ATACTCTACC CAAAAFACT AATCTCAGGG

BssHII  
~~~~~

1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAAATATAG CGCGCAACT AGGATAAATT ATCGCGCGCG  
CGTTAATATG TAAATTATGC GCTATCTTTT GTTTTATATC GCGCGTTTGA TCCTATTAA TAGCGCGCGC

BssHII  
~~~~~

FIG.\_43C

XbaI

ClaI HindIII

1191 GTGTCACTTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAAATTCG CTTATAGTGA  
CACAGTAGAT ACAATGATCT AGCTATTGCA AGATCTCGCC GGCACCTCG AGGTTAAGCG GGTATCACT

BssHII

1261 GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTACAAACGT CGTACTGGG AAAACCCCTGG CGTTACCCAA  
CAGCATAATG CCGCGGAGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC TTTTGGGACC GCAATGGGTT

1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
GAATTAGCGG AACGTCGTGT AGGGGGAAG CCGTCGACCG CATATCGCT TCTCCGGGCG TGGCTAGCGG

1401 CTTCCCAACA GTTGGCGAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGGCGGG  
GAAGGGTTGT CAACGCGTCT GACTTACCGC TTACCTCTGG CCGGACATCG CCGCGTAATT CCGCGCGCCC

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCAGC GCCTAGCGC CCGCTCCCTTT CCGTTCTTTC  
ACACCACCAA TCGCGGTCGC ACTGGCGATG TGAACGGTCT GGGATCGCG GCGGAGGAAA GCGAAAGAAAG

1541 CCTTCCCTTC TCGCCACGTT CCGCGGCTTT CCGCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
GGAAGGAAAG AGCGGTGCAA GCGGCCGAAA GGGCAGTTC GAGATTAGC CCGCGAGGGA AATCCCAAGG

1611 GATTTAGTGC TTACGGCAC CTCGACCCCA AAAAATTGA TTAGGGTGAT GGTTCACGTA GTGGGCCATC  
CTAAATCAGG AATGCCGTG GAGCTGGGCT TTTTGAAC TATCCACTA CCAAGTGCAT CACCCGGTAG

1681 GCCCTGATAG ACGGTTTTC GCCCTTTGAC GTTGGAGTCC ACGTCTTTA ATAGTGGACT CTTGTTCCAA  
CGGACTATC TGCCAAAAG CCGGAAACTG CAACCTCAGG TGCAAGAAAT TATCACCTGA GAACAAGGTT

1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTTTTG ATTATAAGG GATTTGCGG ATTTGGGCTT  
TGACCTTGT GTGAGTTGGG ATAGAGCCAG ATAGAAAAC TAAATATTCC CTAACACGCG TAAAGCCGGA

1821 ATTGGTTAAA AAATGAGCTG ATTTAACAAA AATTTAAGC GAATTTTAA CAAATATTAA CCGTTTACAAT  
TAACCAATTT TTTACTCGAC TAAATGTGTT TAAATTTGCT CTTAAATTTG TTTTATTAAT GCGAATGTTA

FIG. 43D

1891 TTATGTTGGCA CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT TTTCTAAATA CATTCAATA  
AATCCACCGT GAAAGCCCC TTACACGCG CCTTGGGAT AACAAATAA AAAGATTAT GTAAGTTAT

1961 TGTATCCGCT CATGAGACAA TAACCCCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
ACATAGGCGA GTACTCTGTT ATTGGGACTA TTACGAAAT TATTATAACT TTTTCCCTCT CATACTCATA

2031 TCAACATTTC CGTGTGCCCC TTATTCCTTT TTTTGGGCA TTTTGCCTTC CTGTTTTTGC TCACCCAGAA  
AGTTGTAAAG GCACAGCGG AATAAGGGA AAAACGCCGT AAAACGGAAG GACAAAAACG AGTGGGTCTT

2101 ACGTGGTGA AAGTAAAGA TGCTGAAGAT CAGTTGGTG CACGAGTGG TACATCGAA CTGGATCTCA  
TGCAGCACT TTCATTTTCT ACGACTTCTA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT

2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACCT TTTAAAGTTCT  
TGTCGCCATT CTAGGAACTC TCAAAAGCGG GGCTTCTTGC AAAAGGTTAC TACTCGTGAA AATTTCAAGA

2241 GCTATGTGGC GCGGTATTAT CCGGTATTGA CCGCGGGCA GAGCAACTCG GTCGCCGCA ACATATTCT  
CGATACACCG CGCCATAATA GGGCATAACT GCGGCCCGTT CTCGTTGAGC CAGCGGCGTA TGTGATAAGA

2311 CAGATGACT TGGTTGAGTA CTCACCAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAGAGAAAT  
GTCTTACTGA ACCAACTCAT GAGTGGTCAG TGTCTTTTCG TAGAATGCCT ACCGTACTGT CATTCTCTTA

2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACCTGCGGC CAACTTACTT CTGACAACGA TCGGAGGACC  
ATAGGTCAG ACGGTATTGG TACTCACTAT TGTGACGCG GTTGAATGAA GACTGTTGCT AGCTCTCTGG

2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG  
CTTCTCTGAT TGGCGAAAA ACGTGTGTA CCCCCTAGTA CATTGAGCG AACTAGCAAC CCTTGGGCTC

2521 CTGAATGAAG CCATACCAAA CGACGAGCGT GACACACGA TGCCTGTAGC AATGGCAACA ACGTTGGCA  
GACTTACTTC GGTATGGTTT GCTGCTCGCA CTGTGTGCTT ACGGACATCG TTACCGTTGT TGCACGCGT

2591 AACTATTAACT TGGCGAACTA CTTACTCTAG CTTCCCGCA ACAATTAATA GACTGGATGG AGGCGGATAA  
TTGATAATTG ACCGCTTGAT GAATGAGATC GAAGGCGCGT TGTTAATTAT CTGACCTACC TCCGCCATAT

2661 AGTTGCAGGA CCACTTCTGC GCTCGGCCCT TCCGGCTGG TGGTTTATTG CTGATAAATC TGGAGCCCGT  
TCAACGTCCT GGTGAAGACG CGAGCCGGA AGGCCGCG ACCAAATAAC GACTATTAG ACCTCGCGCA

**FIG. 43E**

2731 GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCCGTATC GTAGTTATCT  
CTCGCACCCA GAGCGCCATA GTAACGTCTGT GACCCCGGTC TACCATTCCG GAGGCATAG CATCAATAGA

2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT  
TGTGCTGCCC CTCAGTCCGT TGATACCTAC TTGCTTTATC TGTCTAGCGA CTCTATCCAC GGAGTGACTA

2871 TAAGCATTTG TAACTGTCTAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTTAA  
ATTCTGTAACC ATTGACAGTC TGGTTCAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAATT

2941 TTTAAAAGGA TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCATAAT CCTTTAAAGT GAGTTTTCGT  
AAATTTTCTT AGATCCACTT CTAGGAAAA CTATTAGAGT ACTGGTTTTA GGGAAATGCA CTCAAAAAGCA

3011 TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAAGGATC TTCTTGAGAT CCTTTTTTTC TCGCGGTAAT  
AGGTGACTCG CAGTCTGGG CATCTTTTCT AGTTTCTTAG AAGAACTCTA GGAATAAAG ACGGGCATTA

3081 CTGCTGCTTG CAAACAAAA AACCAACGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT  
GACGACGAAC GTTTGTTTTT TTGGTGGCGA TGGTCGCCAC CAAACAAACG GCCTAGTTCT CGATGGTTGA

3151 CTTTTTCCGA AGGTAAGTGG CTTTCAGCAGA GCGCAGATAC CAAATACGTG CCTTCTAGTG TAGCCGTAGT  
GAAAAAGGCT TCCATTGACC GAAGTCGTCT CGCGTCTATG GTTTATGACA GGAAGATCAC ATCGGCATCA

3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTTACATA CCTCGCTCTG CTAATCCCTG TACCAGTGGC  
ATCCGGTGGT GAAGTTCTTG AGACATCGTG GCGGATGTAT GGAGCGAGAC GATTAGGACA ATGATCACCG

3291 TGCTGCCAGT GGCATTAAGT CGTGTCTTAC CGGTTTGGAC TCAAGACGAT AGTTACCCGA TAAAGGCGCAG  
ACGACGGTCA CCGCTATTCA GCACAGAATG GCCCAACCTG AGTTCTGCTA TCAATGGCCCT ATTCCGCGTC

3361 CGGTCCGGCT GAACGGGGG TTGCTGCACA CAGCCCAAGT TGGAGCGAAC GACCTACACC GAACTGAGAT  
GCCAGCCCGA CTTGCCCCCC AAGCACGTGT GTGCGGTGGA ACTTCGCTTG CTGGATGTGG CTTGACTCTA

3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAAG  
TGGATGTGCG ACTCGATACT CTTTCGCGGT GCGAAGGGCT TCCCTCTTTC CGCCTGTCCA TAGGCCATTC

3501 GGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGTTTCCA GGGGAAACG CCTGGTATCT TTATAGTCTT  
GCCGTCCAG CCTTGTCTTC TCGCGTGTCT CCTGAAGGT CCCCCTTTCG GGACCATAGA AATATCAGGA

**FIG.\_43F**



```

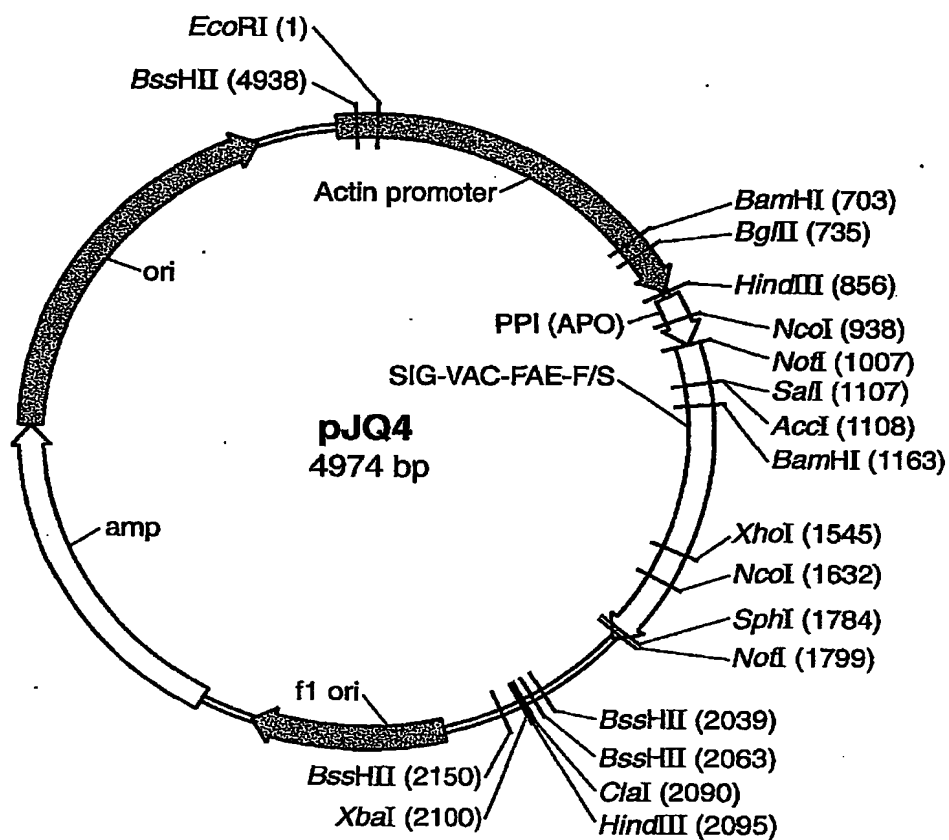
3571 GTGCGGTATC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGTCGTC AGGGGGCGG AGCCTATGGA
 CAGCCCAAAG CGGTGGAGAC TGAACCTGCA GCTAAAACA CTACGAGCAG TCCCCCGCC TCGGATACCT
3641 AAAACGCCAG CAACGCGGCC TTTTACGGT TCCTGGCCCT TTGCTGCTT TTTGCTCACA TGTTCCTTCC
 TTTTGGCGTC GTTGGCGCGG AAAAATGCCA AGGACCGGAA AACGACGGA AACGAGTGT ACAAGAAAGG
3711 TCGGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCGAGC
 ACGCAATAGG GGACTAAGAC ACCTATTGGC ATAATGGCGG AAACCTCACTC GACTATGGCG AGCGGCGTCC
3781 CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCC AATACGCAA A CCGCCTCTCC
 GCTTGCTGGC TCGCGTCGCT CAGTCACTCG CTCCTTCGCC TTCTCGCGGG TTATGCGTTT GCGGAGAGG
3851 CCGCGCGTTG GCCGATTTCAT TAATGCAGCT GGCACGACAG GTTTCCTCGAC TGGAAAGCGG GCAGTGAGCG
 GCGCGCGCAAC CCGCTAAGTA ATTACGTCTGA CCGTGTCTGC CAAAGGCTG ACCTTTCGCC CGTCACCTCGC
3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
 GTTGCCTTAA TTACACTCAA TCGAGTGAGT AATCCGTGGG GTCCGAAATG TGAATACGA AGGCCGAGCA
 BssHII
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTTCACACAG GAAACAGCTA TGACCATGAT TACGCCCAAGC
 TACAACACAC CTTAAACACTC GCCTATTGTT AAGTGTGTC CTTTGTGAT ACTGGTACTA ATGCGGTTCG
 BssHII
                                     ~~~~~
4061  GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGAAT  CCACAATGAA CAATAATAAG ATTAATAATAG
      CGCGTTAATT GGGAGTGATT TCCCTTGTTT TCGACCTTAA GGTGTTACTT GTTATTATT TAAATTTATC
4131  CTTGCCCCCG TTGCAGCGAT GGGTATTTTT TCTAGTAAA  TAAAAGATAA ACTTAGACTC AAAACATTTA
      GAACGGGGC AACGTGCTA CCCATAAAA AGATCATTTT ATTTCTATT TGAATCTGAG TTTTGTAAAT
4201  CAAAACAAC CCCTAAAGTC CTAAAGCCCA AAGTCTATG  CACGATCCAT AGCAAGCCCA GCCCAACCCA
      GTTTTGTGTT GGGATTTCAG GATTTCGGGT TTCAAGATAC GTGCTAGGTA TCGTTCGGGT CCGGTTGGGT

```

**FIG. 43G**

4271 ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTAGT CACCGTAGT  
 TGGGTGGGT TGGGTGGGT CACGTCGGT CACGTCGGT CAGAGGTGG GGGCGTGATA GTGGCACTCA  
 4341 TGTCGGACC ACCGCACGTC TCGCAGCCAA AAAAAAATA AGAAGAAAA AAAAGAAAA GAAAAACAGC  
 ACAGGCGTGG TGGCGTGCAG AGCGTCGGT TTTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTTCTG  
 4411 AGGTGGTCC GGTCTGTGG GGGCGGAAA GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC  
 TCCACCCAGG CCCAGCACCC CCGGCTTTT CGCTCTTCTT AGCGCTCGTC GCTGCTCCGG GCCGGGAGGG  
 4481 TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCCC TCTCTCCA TCCCCCAAC  
 AGCGAAGGT TTCTTTCGG GGGGTAGCG TGATATATGT ATGGGGGGG AGAGGAGGT AGGGGGTTG  
 4551 CCTACACCA CCACCACCAC CACCTCTCC CCCCTCGCT CCGGACGAC AGCTCTCTCC CCTCCCCCT  
 GGATGGTGGT GGTGGTGGT GTGGAGGAG GGGGAGCGAC GGCCTGCTGC TCGAGGAGGG GGGAGGGGA  
 4621 CCGCCGCCGC CGGTAACCAC CCGCCCCCT TCTCTTCT TCTCTCTT TTTTTTCT CTCGTCCTG  
 GCGGCGGCG GCCATTGGT GGGCGGGAG AGGAGAAAG AAGAGGCAA AAAAAAGCA GAGCCAGAC  
 4691 ATCTTTGGC TTGGTAGTTT GGGTGGCGA GAGCGGCTC GTCGCCAGA TCGGTGCGG GAGGGGGCG  
 TAGAAACCGG AACCATCAAA CCCACCCGCT CTCGCCGAG CAGCGGCT AGCCACGCG CCTCCCCGCC  
 BglII BamHI  
 4761 GATCTCGCG CTGGCGTCTC CGGCGGTGAG TCGGCCCGGA TCCTCGCGGG GAATGGGGT CTCGGATGTA  
 CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCTT AGGAGCGCCC CTTACCCCGA GAGCCTACAT  
 BglII  
 4831 GATCTCTCTT CTTTCTTCTT TTTGTGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTCTT  
 CTAGAGAAA GAAAGAAA AAACCAATC TTAACCTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAAGAA  
 4901 TTCAATGATTT GTGACAAATG CAGCCTCGT CCGAGCTTTT TTGTAGGTAG  
 AAGTACTAAA CACTGTTTAC GTCGAGCAC GCCTCGAAA AACATCCATC

FIG. 43H

**FIG. 44A**

**FIG. 44B**

```

771  AATCCCTCAG CATTTGTCAT CGGTAGTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTCCGGAGC
    TTAGGGAGTC GTAAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG

    HindIII
    ~~~~~
841 TTTTTTTGTAG GTAGAAGCTT ACMATGCMCG TGCACAAGGA GGTSAACTTC GTSGCCTACC TCCTGATCGT
 AAAAAACATC CATCTTCGAA TGTACCKGC ACGTGTCCT CCASTTGAAG CASC GGATGG AGGACTAGCA

 NcoI
    ~~~~~
911  SCTCGGCTC CTCTTGCTCG TSTCCGCCAT GGAGCAGTG GACGCCAAGG CCTGCACCCK CGAGTGCAGC
    SGAGCCGGAG GAGAACGAGC ASAGCGGTA CTCTGCGAC CTGCGGTTCC GGACGTGGGM GCTCACGCCG

    NotI
    ~~~~~
981 AACCTCGGCT TCGGCATCTG CCCGGCGGCC GCCTCCAGC AGGCATCTC CGAAGACCTC TACAGCCGTT
 TTGAGCCGA AGCCGTAGAC GGGCCGCCGG CGGAGGTGG TCCCGTAGAG GCTTCTGGAG ATGTCGGCAA

 SalI
    ~~~~~
    AcoI
    ~~~~~
1051 TAGTCGAAAT GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA
 ATCAGCTTTA CCGGTGATAG AGGTTTCGAC GGATGCGGT GGACACGTTG TAAGGCAGCT GATAATAGTT

 BamHI
    ~~~~~
1121 GGGAGAGAAA ATTTACAATT CTCAAACTGA CAATTAAGGA TGGATCCTCC GCGACGACAG CAGCAAGAA
    CCCCTCTTTT TAAATGTTAA GAGTTTGACT GTAATTGCCT ACCTAGGAGG CGCTGCTGTC GTCGTTCTTT

1191 ATAATCACCG TCTTCGCTGG CACTGGTAGT GATACGAATC TACAACTGTA TACTAACTAC ACCCTCACGC
    TATTAGTGGC AGAAGGCACC GTGACCATCA CTATGCTTAG ATGTTGAGCT ATGATTGATG TGGGAGTGGC

1261 CTTTCGACAC CCTACCACAA TGCACCGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT
    GAAAGCTGTG GGATGGTGT ACCTTGCCAA CACTTCATGT GCCACCTATA ATATAACCTA CCCAGAGGCA

```

**FIG.\_44C**

```

1331 CCAGGACCAA GTCGAGTCGC TTGTCAAACA GCAGGTAGC CAGTATCCGG ACTACGGCT GACCGTGACC
    GGTCCCTGGTT CAGCTCAGCG AACAGTTTGT CGTCCAATCG GTCATAGGCC TGATGCGCGA CTGGCACTGG

1401 GGCCACKCCC TCGGCGCCTC CCTGGCGGCA CTCACTGCCC CCCAGCTGTC TCGGACATAC GACAAACATCC
    CCGGTGCGGG AGCCGCGGAG GGACCGCCGT GAGTGACGGC GGGTCGACAG ACGCTGTATG CTGTTGTAGG

1471 GCCTGTACAC CTTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTGCGGTGC TACATGAACG ATGCCCTTCCA
    CGGACATGTG GAAGCCGCTT GCGCGCTGCG CGTTAGTCCG GAAGCGCAGC ATGTACTTGC TACGGAAGGT

    XhoI
    ~~~~~
1541 AGCCTCGAGC CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC
 TCGGAGCTCG GGTCTATGCT GCGTCATAAA GCGCCAGTGA GTACGGTTGC TGCCGTAGGG TTTGGACGGG

 NcoI
    ~~~~~
1611 CCGGTGGAGC AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAAC
    GGCCACCCTCG TCCCCATGCG GGTACCGCCA CATCTCATGA CCTCGCACT AGGAATGTCTG CCGGTCTTGT

1681 CATTTGTCG CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCCCA
    GTAAACAGAC GTGACCCCTTA CTTACAGTCA CGACACTCCG GGTCCCGCT GTCCACACT TATTACGCGT

    SphI
    ~~~~~
1751 CACGACTTAT TTTGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTGCGGG CCGCGGAAC CACTGAAGGA
 GTGCTGAATA AAACCCCTACT GCTCGCCGCG TACGTGGACC GGCAGCGCC GGCCTTTG GTGACTTCTT

1821 TGAGCTGTAA AGAAGCAGAT CGTTCAAACA TTTGGCAATA AGTTTCTTGA AGATTGAATC CTGTTGCCGG
 ACTCGACATT TCTTCGTCTA GCAAGTTTGT AAACCGTTAT TTCAAAGAAT TCTAACTTAG GACAACGGCC

1891 TCTTGGGATG ATTATCATAT AATTTCGTGT GAATTACGTT AAGCATGTAA TAATTAACAT GTAATGCATG
 AGAACGCTAC TAATAGTATA TTAAGACAA CTTAATGCAA TTCGTACATT ATTAATTGTA CATTACGTAC

1961 ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCCGC AATTATACAT TTAATACGCG ATAGAAAACA
 TGCAATAAAT ACTCTACCCA AAAATACTAA TCTCAGGGCG TTAATATGTA AATTATGCGC TATCTTTTGT

```

FIG.-44D

|      | BssHII<br>~~~~~                                                                                                                                              | BssHII<br>~~~~~ | ClaI HindIII<br>~~~~~ | XbaI<br>~~ |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|-----------------------|------------|
| 2031 | AAATATAGCG CGCAAACTAG GATAAATTAT CGCGCGCGGT GTCATCTATG TTACTAGATC GATAAGCTTC<br>TTTATATCGC GCGTTTGATC CTATTTAATA GCGCGCGCCA CAGTAGATAC AATGATCTAG CTATTCGAAG |                 |                       |            |
|      | XbaI<br>~~~~~                                                                                                                                                | BssHII<br>~~~~~ |                       |            |
| 2101 | TAGAGCGGCC GGTGGAGCTC CAATTCGCC TATAGTGAGT CGTATTACGC GCGCTCACAG GCCGTCGTTT<br>ATCTCGCCGG CCACCTCGAG GTTAAGCGGG ATATCACTCA GCATAATGCG CGCGAGTGAC CGGCAGCAAA  |                 |                       |            |
| 2171 | TACAACGTCG TGACTGGGAA AACCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC CCCCTTTTCG<br>ATGTTGCAGC ACTGACCCTT TTGGGACCGC AATGGGTTGA ATTAGCGGAA CGTCGTGTAG GGGGAAAGCG  |                 |                       |            |
| 2241 | CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT TCGGCAGCCT GAATGGCGAA<br>GTCGACCGCA TTATCGCTTC TCCGGGCGTG GCTAGCGGGA AGGTTGTCA ACGCGTCGGA CTTACCGCTT  |                 |                       |            |
| 2311 | TGGGACGCGC CCTGTAGCGG CGCATTAAGC GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG ACCGCTACAC<br>ACCTGCGCG GGAATCGCC GCGTAATTCG GCGCGCCAC ACCACCAATG CGCGTCGCAC TGGCGATGCG    |                 |                       |            |
| 2381 | TTGCCAGCGC CCTAGCGCCC GCTCCCTTTCG CTTTCTTCCC TTCTTTTTCG GCCACGTTTCG CCGCTTTTCC<br>AACGGTCCG GGAATCGCG GAGGGAAGC GAAAGAAGGG AAGGAAGAG CGGTGCAAGC GGCAGAAAGG   |                 |                       |            |
| 2451 | CCGTCAGCT CTAAATCGGG GGTCCCTTT AGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA<br>GGCAGTTTCA GATTAGCCC CCGAGGGA TCCCAAGCT AATCACGAA ATGCCGTGGA GTTGGGGTTT         |                 |                       |            |
| 2521 | AAACTTGATT AGGGTGATGG TTCACGTAGT GGGCATCGC CCTGATAGAC GGTTTTTCG CCTTTGACGT<br>TTTGAACATA TCCCACTACC AAGTGCAATCA CCGGTAGCG GGAATATCTG CCAAAAAGCG GGAATACTGCA  |                 |                       |            |
| 2591 | TGGAGTCCAC GTTCTTTAAT AGTGGACTCT TGTTCCAAAC TGAACAACA CTCAACCCCTA TCTCGGTCTA<br>ACCTCAGGTG CAAGAAATTA TCACCTGAGA ACAAGGTTG ACCTTGTGT GAGTTGGGAT AGAGCCAGAT   |                 |                       |            |
| 2661 | TTCTTTTGAT TTATAAGGGA TTTTGGCGAT TTCGGCTAT TGGTTAAAA ATGAGCTGAT TTAACAAAAA<br>AAGAAAACATA AATATCCCT AAAACGGCTA AAGCGGATA ACCAATTTT TACTCGACTA AATGTGTTTT     |                 |                       |            |

FIG.\_44E

2731 TTATAACGGCA ATTTTAACAA AATATTAAACG CTTACAAATTT AGGTGGCACT TTTCGGGGAA ATGTGCGCGG  
 AAATGCGCT TAAAATTGTT TTATAATTGC GAATGTTAAA TCACACCGTGA AAAGCCCCCTT TACACGCGCC  
 2801 AACCCCTAAT TGTATTATTT TCTAAATATA TCTCAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA  
 TTGGGGATAA ACAATAAAA AGATTATGT AAGTTATATC ATAGGCGAGT ACTCTGTTAT TGGGACTATT  
 2871 ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCG TGTCGCCCTT ATTCCCCTTT  
 TACGAAGTTA TTATAACTTT TTCTTCTCA TACTCATAG TTGTAAAGGC ACAGCGGGA TAAGGGAAAA  
 2941 TTGCGGCATT TTGCTTCTT TTGCTTCTT GTTTTTCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA  
 AACGCCATA AACGGAAGGA CAAAACGAG TGGGTCTTTG CGACCATTCT CATTTCTAC GACTTCTAGT  
 3011 GTTGGTGCA CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAAG TCCTTGAGAG TTTTCGCCCC  
 CAACCCACGT GCTCACCCAA TGTAGCTTGA CCTAGATTG TCGCCATTCT AGGAACCTCTC AAAAGCGGGG  
 3081 GAAGAACGTT TTCCAATGAT GAGCATTCTT AAAGTTCTGC TATGTGGCGC GGTATTATCC CGTATTGACG  
 CTTCTTGCA AAGGTTACTA CTCGTGAAA TTTCAAGACG ATACACCGCG CCATAATAGG GCATAACTGC  
 3151 CCGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC  
 GGGCCGTTCT CGTTGAGCCA GCGCGGTATG TGATAAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG  
 3221 AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGATTTA TGCAGTCTG CCATAACCAT GAGTGATTAAC  
 TCTTTTTCGTA GAATGCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGTA CTCACTATTG  
 3291 ACTGCGGCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG  
 TGACGCCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAAAC GTGTTGTACC  
 3361 GGGATCATGT AACTCGCTT GATCGTTGGG AACCGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA  
 CCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCGG TATGGTTTGC TGCTCGCACT  
 3431 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGGCGAAA CTATTAACTG GCGAACTACT TACTCTAGCT  
 GTGTGCTAC GGACATCGTT ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA  
 3501 TCCGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC  
 AGGGCCGTTG TTAATTATCT GACCTACCTC CGCCTATTTT AACGTCCTGG TGAAGACGCG AGCCGGGAG

**FIG.-44F**



3571 CGGCTGGCTG GTTATTTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT  
 GCCGACCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCAGA GCGCCATAGT AACGTCGTGA  
 3641 GGGGCCAGAT GGTAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA  
 CCCCCTCTA CCATTCGGGA GGGCATAGCA TCAATAGATG TGCTGCCCTT CAGTCCGTG ATACCTACTT  
 3711 CGAAATAGAC AGATCGCTGA GATAGGTGCC TCACTGATTA AGCAATTGTA ACTGTGAGAC CAAGTTTACT  
 GCITTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA  
 3781 CATATATACT TTAGATTGAT TTAATACTTC ATTTTAAAT TAAAAGGATC TAGGTGAAGA TCCTTTTGA  
 GTATATATGA AATCTAATA AATTTTGAAG TAAAAATTAA ATTTCCCTAG ATCCACTTCT AGGAAAAACT  
 3851 TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTTCGTT CACTGAGCGT CAGACCCCGT AGAAAAGATC  
 ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGCA TCTTTTCTAG  
 3921 AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGGTAATCT GCTGCTTGCA AACAAAAA CCACCGCTAC  
 TTTCCCTAGAA GAATCTTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG  
 3991 CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC  
 GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGGTTGAGA AAAAGGCTTC CATTCACCGA AGTCGTCTCG  
 4061 GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCACCACT TCAAGAATC TGTAGCACCG  
 CGTCTATGGT TTATGACAGG AAGATCACAT CGGCATCAAT CCGGTGGTGA AGTTCTTGAG ACATCGTGCC  
 4131 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCC TGTCTTACC  
 GGATGTATGG AGCGAGACGA TTAGGACAAT GGTACCCGAC GACGGTCACC GCTATTTCAGC ACAGAAATGGC  
 4201 GGTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGTT CGTGACACA  
 CCAACCTGAG TTCTGCTATC AATGGCCTAT TCCGCGTCCG CAGCCCGACT TGCCCCCAA GCACGTGTGT  
 4271 GCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG  
 CCGGTCCGAC CTCGCTTGCT GGATGTGGCT TGACTCTATG GATGTCCAC TCGATACTCT TTCGCGGTGC  
 4341 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCCG AACAGGAGAG CGCACGAGGG  
 GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTCCG CGTCCAGCC TTGTCTCTCT GCCTGCTCCC

**FIG. 44G**

4411 AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCTGTG CGGTTTTCGC CACCTCTGAC TTGAGCGTTCG  
 TCGAAGGTCC CCTTTGCGG ACCATAGAAA TATCAGGACA GCCCAAAGCG GTGGAGACTG AACTCGCAGC  
 4481 ATTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGA AAACCCAGCA ACGCGSCCTT TTTACGGTTC  
 TAAAAACACT ACGAGCAGTC CCCCCGCCTC GGATACCTTT TTGCGGTCTGT TCGCGCGGAA AATGCGCAAG  
 4551 CTGGCCCTTT GCTGGCCCTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCGTG GATAAACCGTA  
 GACCGGAAAA CGACCGGAAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT  
 4621 TTACCGCCCTT TGAGTGAGCT GATACCGCTC GCGGAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA  
 AATGGCGGAA ACTCACTCGA CTATGGCGAG CGGCGTCGCG TTGCTGGCTC GCGTCGCTCA GTCACTCGCT  
 4691 GGAAGCGGAA GAGCGCCCAA TACGCAAAACC GCTCTTCCCC GCGCGTTGGC CGATTCAATTA ATGCAGCTGG  
 CCTTCGCCCTT CTCGCGGGTT ATGCGTTTGG CGGAGAGGGG CGCGCAACCG GCTAAGTAAT TACGTCGACC  
 4761 CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTGAGCGCA ACGCAATTA TGTGAGTTAG CTCACCTCAT  
 GTGCTGTCCA AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TCGGTAAAT ACACFCAATC GAGTGAGTAA  
 4831 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAAT  
 TCCGTGGGGT CCGAAATGTG AATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTATA  
  
 4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAACC CTCACTAAG GGAACAAAAG  
 AGTGTGTCCT TTGTCGATAC TGGTACTAAT GCGGTTCCGG CGTTAATTGG GAGTGATTTT CCTTGTTC

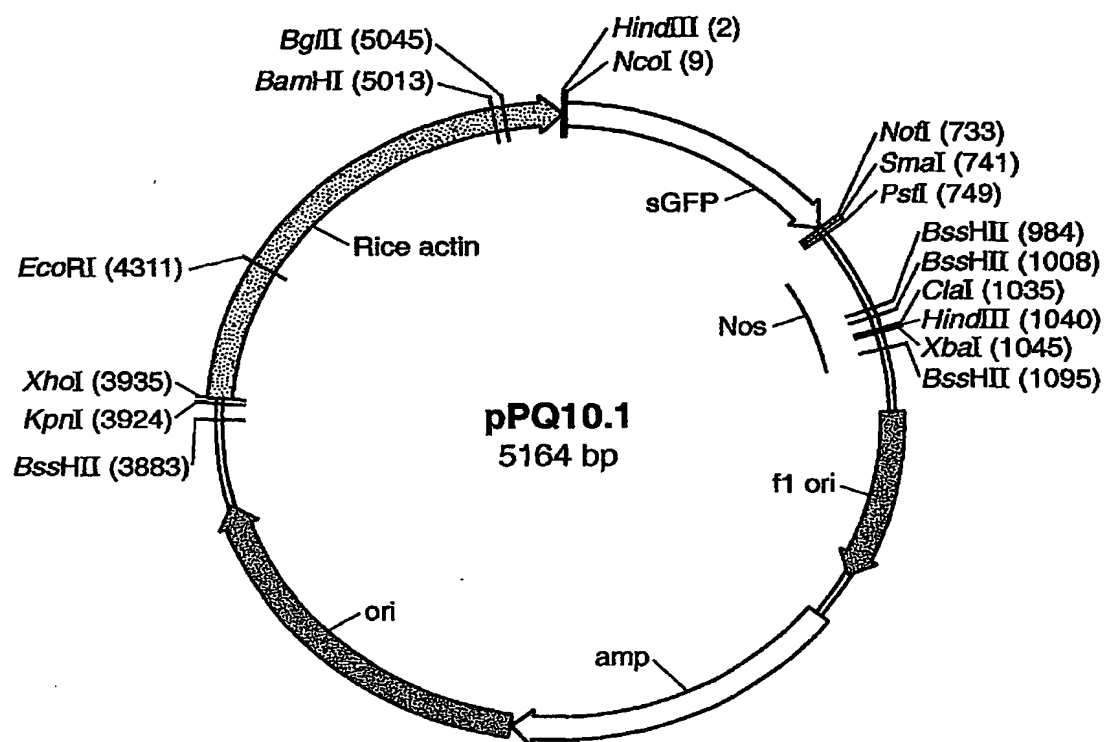
BssHII

\*\*\*\*\*

EcoR

CTGG  
GACC

FIG. 44H

**FIG.\_45A**

## Sequence for pPQ10.1

HindIII NcoI  
-----  
1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGGTGGTGCC  
CATCCTGGTC GAGCTGGACG  
TTCGAATGGT ACCACTCGTT CCGCTCCTC GACAAGTGGC CCCACCACGG  
GTAGGACCAG CTCGACCTGC  
  
71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT  
GCCACCTACG GCAAGCTGAC  
CGCTGCACTT GCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA  
CGGTGGATGC CGTTCGACTG  
  
141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCACCC  
TCGTGACCAC CTTACCTTAC  
GGACTTCAAG TAGACGTGGT GGCCTTCGA CGGGCACGGG ACCGGGTGGG  
AGCACTGGTG GAAGTGGATG  
  
211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGGACTT  
CTTCAAGTCC GCCATGCCCG  
CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA  
GAAGTTCAGG CGGTACGGGC  
  
281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC  
AAGACCCGCG CCGAGGTGAA  
TTCCGATGCA GGTCTCTCGG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG  
TTCTGGGCGC GGCTCCACTT  
  
351 GTTCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT  
TCAAGGAGGA CGGCAACATC  
CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA  
AGTTCCTCCT GCCGTTGTAG  
  
421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT  
GGCCGACAAG CAGAAGAACG  
GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA  
CCGGCTGTTT GTCTTCTTGC  
  
491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG  
CAGCTCGCCG ACCACTACCA  
CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTGCGAC  
GTCGAGCGGC TGGTGATGGT  
  
561 GCAGAACACC CCCATCGGCG ACGGCCCGT GCTGCTGCCC GACAACCACT  
ACCTGAGCAC CCACTCCGCC  
CGTCTGTGG GGGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA  
TGGACTCGTG GGTACGGCGG  
  
631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT  
CGTGACCGCC GCCTGGATCA

Fig. 45B

CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGGT TGTCAACGCG  
TCGGACTTAC CGCTTACCCCT

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA  
GCGTGACCGC TACACTTGCC  
GCGCGGGACA TCGCCGCGTA ATTGCGCGCG CCCACACCAC CAATGCGCGT  
CGCACTGGCG ATGTGAACGG

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC  
GTTGCGCCGC TTTCCCGTC  
TCGCGGGATC GCGGGCGAGG AAAGCGAAAG AAGGGAAGGA AAGAGCGGTG  
CAAGCGGCCG AAAGGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTAGGGT TCCGATTAG TGCTTTACGG  
CACCTCGACC CCAAAAACT  
TTCGAGATT AGCCCCGAG GGAAATCCCA AGGCTAAATC ACGAAATGCC  
GTGGAGCTGG GGTTTTTTGA

1471 TGATTAGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA TAGACGGTTT  
TTCGCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACC CGG TAGCGGGACT ATCTGCCAAA  
AAGCGGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTTGTT CAACTGGAA CAACACTCAA  
CCCTATCTCG GTCTATTCTT  
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT  
GGGATAGAGC CAGATAAGAA

1611 TTGATTATA AGGGATTTTG CCGATTTGCG CCTATTGGTT AAAAAATGAG  
CTGATTAAAC AAAAATTTAA  
AACTAAATAT TCCCTAAAC GGCTAAAGCC GGATAACCA TTTTTTACTC  
GACTAAATTG TTTTAAATT

1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTTAGGTG GCACTTTTCG  
GGGAAATGTG CGCGGAACCC  
GCGCTTAAAA TTGTTTTATA ATTGCGAATG TTAAATCCAC CGTGAAAAGC  
CCCTTTACAC GCGCCTTGGG

1751 CTATTTGTTT ATTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA  
CAATAACCCCT GATAAATGCT  
GATAAACAAA TAAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT  
GTTATTGGGA CTATTTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCG  
CCCTTATTCC CTTTTTTGCG  
AGTTATTATA ACTTTTTCCT TCTCATCTC ATAAGTTGTA AAGGCACAGC  
GGGAATAAGG GAAAAAACGC

1891 GCATTTTGCC TTCTGTTTT TGCTCACCCA GAAACGCTGG TGAAAGTAA  
AGATGCTGAA GATCAGTTGG  
CGTAAACGG AAGGACAAA ACGAGTGGGT CTTTGCGACC ACTTTCATTT  
TCTACGACTT CTAGTCAACC

Fig. 45C

1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT  
GAGAGTTTTC GCCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTGCGC ATTCTAGGAA  
CTCTCAAAAG CGGGGCTTCT

2031 ACGTTTTCCTA ATGATGAGCA CTTTAAAGT TCTGCTATGT GCGCGGGTAT  
TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAAATTTCA AGACGATACA CCGCGCCATA  
ATAGGGCATA ACTGCGGCCC

2101 CAAGAGCAAC TCGGTGCGCG CATACACTAT TCTCAGAATG ACTTGTTTGA  
GTACTCACCA GTCACAGAAA  
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT  
CATGAGTGGT CAGTGTCTTT

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA  
ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT  
TGGTACTCAC TATTGTGACG

2241 GGCCAACCTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT  
TTTTGCACAA CATGGGGGAT  
CCGTTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA  
AAAACGTGTT GTACCCCTA

2311 CATGTAACCT GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC  
AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAAGTAGC AACCCTTGGC CTCGACTTAC TTCGGTATGG  
TTTGCTGCTC GCACTGTGGT

2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA  
CTACTTACTC TAGCTTCCCG  
GCTACGGACA TCGTTACCGT TGTTGCAACG CGTTTGATAA TTGACCGCTT  
GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC  
TGCGCTCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCCT ATTTCAACGT CTTGGTGAAG  
ACCGGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG  
TATCATTGCA GCACTGGGGC  
CCGACCAAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC  
ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG  
GCAACTATGG ATGAACGAAA  
GTCTACCAT TCGGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC  
CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACTGT  
CAGACCAAGT TTAATCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA  
GTCTGGTTCA AATGAGTATA

Fig. 45D

2731 ATACTTTTGA TTGATTTAAA ACTTCAATTTT TAATTTAAAA GGATCTAGGT GAAGATCCTTT TTTGATAATC  
TATGAAATCT AACTAAATTT TGAAGTAAAA ATTAATTTT CCTAGATCCA CTTCCTAGGAA AAACATATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CGTTCACCTG AGGTCAGAC CCGTAGAAAA AGATCAAAGG  
AGTACTGGTT TTAGGGAAAT GCACCTAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCTTTT TTCTGGCGGT AATCTGCTGC TTGCAAAACAA AAAAACACC GCTACCAGCG  
TAGAAGAACT CTAGGAAAAA AGACGGCA TTAGACGACG AACGTTTGT TTTTGGTGG CGATGGTCCG

2941 GTGGTTTGT TTCCCGGATCA AGACTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA  
CACCAAAACAA ACGGCCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAGTCG TCTCGCGTCT

3011 TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTTCAAG AACTCTGTAG CACCGCCTAC  
ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTACCAGT GGTGCTGCC AGTGCGATA AGTCGTGTCT TACCGGGTTG  
TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCAAC

3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTGG GGTTCGTGC ACACAGCCCA  
CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTCCGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCCTGAGCTA TGAGAAAGCG CCACGCTTCC  
CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT  
GCTTCCCTCT TTCCCGCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGT CTTCTCGCTG CTCCCCTCGAA

3361 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT  
GGTCCCCCTT TCGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGA GACTGAATC GCAGCTAAA

3431 TGTGATGCTC GTCAGGGGG CGGAGCCTAT GGA AAAACGC CAGCAACCG GCCTTTTAC GGTTCCTGGC  
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCTGTGGC CGGAAAAATG CCAAGGACCG

**FIG. 45E**

```

3501 CTTTTGCTGG CTTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATTA CCGTATTACC
 GAAACGACC GGAACGAG TGTACAAGAA AGGACGCAAT AGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAG
 CGGAAACATCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGTCTGCGTC GCTCAGTCAC TCGCTCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCCGCGG TGGCCGATT CATTAATGCA GCTGGCACGA
 GCCTTCTCGC GGGTATGCG TTTGGCGGAG AGGGCGCGC AACCGCTAA GTAAATTACGT CGACCGTGT

3711 CAGGTTTCCC GACTGGAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAG TCATTAGGCA
 GTCCAAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTTACA
 GGGGTCCGAA ATGTGAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCATT GTTAAAGTGT

 BssHII
3851 CAGGAACACG CTATGACCAT GATTACGCCA AGCGCGCAAT TAACCTCAGT TAAAGGGAAC AAGAGCTGGG
 GTCCTTTGTC GATACTGGTA CTAATGCGGT TCGCGCGTTA ATTGGGAGTG ATTTCCCTTG TTTTCGACCC

 KpnI
 XhoI
3921 TACCGGGCCC CCCCTCGAGG TCATTTCATAT GCTTGAGAAG AGAGTCGGGA TAGTCCAAA TAAACAAAG
 ATGGCCCGGG GGGGAGCTCC AGTAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTTCTTTC

3991 GTAAGATTAC CTGGTCAAAA GTGAAAACAT CAGTTAAAAG GTGGTATAAG TAAATATCG GTATATAAAG
 CATTCATAAG GACCAGTTTT CACTTTTGTG GTCAATTTTC CACCATATTC ATTTTATAGC CATTAATTTTC

4061 GTGGCCCAAA GTGAAATTTA CTCTTTCTTA CTATTAAAA AATTGAGGAT GTTTTGTGCG TACTTTTGATA
 CACCGGGTTT CACTTTAAAT GAGAAAAGAT GATAATATTT TTAACCTCTA CAAAACAGCC ATGAACATAT

```

FIG. 45F



4131 CGTCATTTTT GTATGAATTG GTTTTTTAAGT TTATTCGCGA TTTCGAAATG CATATCTGTA TTTGAGTCGG  
GCAGTAAAAA CATACTTAAC CAAAAATCA AATAAGCGCT AAACCTTTAC GTATAGACAT AAACTCAGCC

4201 TTTTAAAGTT CGTTGCTTTT GTAAATACAG AGGATTGTG ATAAGAAATA TCTTTAAAA ACCCATATGC  
AAAAATTCAA GCAACGAAAA CATTTATGTC TCCCTAAACA TATTCCTTAT AGAAATTTTT TGGGTATACG

EgRI  
\*\*\*\*\*

4271 TAATTGACA TAATTTTGA GAAAAATATA TATTCAGCG TATCCACAA TGAACAATA TAAGATTAA  
ATAAACTGT ATAAAAACT CTTTTTATAT ATAAGTCGC TTAAGGTGTT ACTTGTATT ATCTTAATTT

4341 ATAGCTTGCC CCCCTTGCAG CGATGGGTAT TTTTTCTAGT AAAATAAAG ATAACTTAG ACTCAAAACA  
TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA TTTTATTTC TATTGAATC TGAGTTTTGT

4411 TTTACAAAAA CAACCCCTAA AGTCTTAAAG CCCAAAGTGC TATGCACGAT CCATAGCAAG CCCAGCCCCA  
AAATGTTTTT GTTGGGANT TCAGGATTC GGGTTTCACG ATACGTGCTA GGTATCGTTC GGGTCGGGTT

4481 CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA AATAGTCTCC ACCCCGGCA CTATCACCGT  
GGGTTGGGTT GGGTTGGGTG GGGTCACGTC GGTGACCGT TTATCAGAG TGGGGGCCGT GATAGTGGCA

4551 GAGTTGTCG CACCACGCA CGTCTCGCAG CCAAAAAA AAAAGAAAG AAAAAAGA AAAAGAAAA  
CTCAACAGG GTGTGGCGT GCAGAGCGTC GGTTTTTTTT TTTTCTTTC TTTTCTTTT

4621 CAGCAGGTGG GTCCGGGTG TGGGGGCCGG AAAAGCGAG AGGATCGCGA GCAGCGACGA GGGCCGGCCC  
GTCGTCCACC CAGGCCAGC ACCCCGGCC TTTTTCGCTCC TCTAGCGCT CGTCGCTGCT CCGGGCCGGG

4691 TCCCTCCGCT TCCAAAGAA CGCCCCCAT CGCCACTATA TACATACCCC CCCCTCTCT CCCATCCCCC  
AGGAGGGCGA AGGTTCTTTT GCGGGGGGTA GCGGTGATAT ATGTATGGGG GGGGAGAGGA GGGTAGGGGG

4761 CAACCTACC ACCACACCA CCACCACCTC CTCCCCCTC GTGCCGGAC GACGAGCTCC TCCCCCTCC  
GTTGGGATGG TGTGTGTGTT GGTGTGGAG GAGGGGGAG CGACGGCCTG CTGCTCGAG AGGGGGAGG

4831 CCTCCGCCG CCGCCGGTAA CCACCCGCC CCTCTCTCT TTTCTCTCC GTTTTTTTTT TCGTCTCGGT  
GGGAGGCCGC GCGGCCATT GGTGGGCCG GGAGAGGAGA AAGAAAGAG CAAAAAAA AGCAGAGCCA

**FIG. 45G**

4901 CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG CTTCTGTCGCC CAGATCGGTG CGCGGGAGGG  
GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCCTCCCTCCC

BamHI  
\*\*\*\*\*

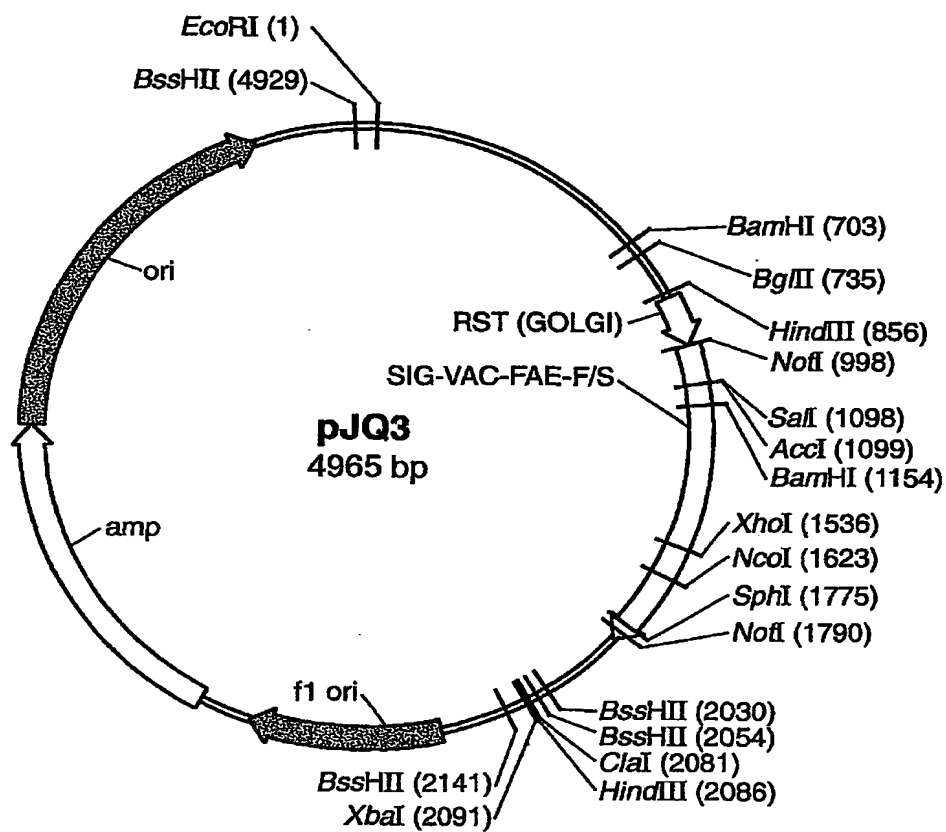
4971 GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGATCCTCG CCGGGAATGG GGCTCTCGGA  
CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCCG GCCTAGGAGC GCCCTTACC CCGAGAGCCT

BglII  
\*\*\*\*\*

5041 TGATAGATCTT CTTTCTTTCT TCTTTTGTG GTAGAAATTG AATCCCTCAG CATTGTTCAT CGGTAGTTTT  
ACATCTAGAA GAAAGAAAG AGAAAAACAC CATCTTAAAC TTAGGGAGTC GTAACAAGTA GCCATCAAAA

5111 TCTTTTCATG ATTGTGACA AATGCAGCCT CGTGGGAGC TTTTGTGTAG GTAG  
AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG AAAAACATC CATC

FIG.\_45H

**FIG. 46A**

EcoRI  
1 AATCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCGTTGACAG CGATGGGTAT TTTTCTTAGT  
TTAAGGTGTT ACTTGTTATT ATCTTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA

71 AAAATAAAG ATAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTCCATAAG CCCAAAGTGC  
TTTTATTTC TATTGAATC TGAGTTTGT AAATGTTTTT GTTGGGATT TCAGGATTTT GGGTTTCACG

141 TATGCACGAT CCATAGCAAG CCCAGCCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA  
ATACGTGCTA GGTATCGTTC GGGTCGGGTT GGGTTGGGTT GGGTCACGTC GGTTCACCGT

211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCAACGCA CGTCTGCGAG CCAAAAAA  
TTATCAGAGG TGGGGCCCGT GATAGTGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTTT

281 AAAAGAAG AAAAAGA AAAAAGAAA CAGCAGGTGG GTCCGGGTGG TGGGGGCCGG AAAAGCGAGG  
TTTTCTTTC TTTTCTTCT TTTTCTTCTT GTCTCTCAC CAGGCCCAGC ACCCCCGGCC TTTTCGCTCC

351 AGGATCGCGA GCAGCGACGA GGCCCGGCC TCCCTCCGCT TCCAAAGAAA CGCCCCCAT CGCCACTATA  
TCCTAGCGCT CGTCGTGCT CCGGGCCGG AGGAGGCGA AGGTTCTTTT GCGGGGGTA GCGTGTATAT

421 TACATACCC CCCCTCTCT CCATFCCC CAACCTFACC ACCACACCA CCACCACTC CTCCTCCCTC  
ATGTATGGG GGGAGAGGA GGTAGGGGG GTTGGGATGG TGGTGTGGT GGTGTGGAG GAGGGGGAG

491 GGTGCCGAC GACGAGCTCC TCCCCCTNCC CCTCCGCGG CCGCCGGTAA CCACCCCGCC CCTCTCTCT  
CGACGGCTG CTGCTCGAG AGGGGGGAG GGGAGGCGGC GCGGCCCATT GGTGGGGCGG GAGAGGAGA

561 TTCCTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG  
AAGAAAGAG CAAAAAAGAA AGCAGAGCCA GAGTAGAAA CCGGAACCAT CANACCCACC CGCTCTCACC

631 CTTCTGCTCC CAGATCGGTG GCGGGAGGG GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC  
GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC CCGCCTAGAG CCGCGACCGC AGAGGCCCGC ACTCAGCCCG

BamHI  
701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGATGATCTT CTTTCTTCTT TCTTTTTTGT GTAGAATTGG  
GCCTAGGAGC GCCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAAC

BglII

**FIG. 46B**

```

771 AATCCCTCAG CATGTTTCAT CGGTAGTTTT TCTTTTCATG ATTGTGACA AATGCAGCCTT CGTCCGGAGC
TTAGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCTCG

HindIII
~~~~~
841 TTTTTTGTAG GTAGAAGCTT ACCATGATCC ACACCAACCT CAAAAGAGG TTCTCCCTCTT TCATCCTCGT
AAAAACATC CATCTTCGAA TGGTACTAGG TGTGGTTGGA GTTTTCTTTC AAGAGGGAGA AGTAGGAGCA

911 CTTCTCCTC TCGCCGTTGA TCTGGGTGTG GAAGAAGGGC TCCGACTACG AGGCCCTCAC CCTCCAAGCC
GAAGGAGGAG AAGCGGCACT AGACGCACAC CTTCTTCCCG AGGCTGATGC TCCGGGAGTG GGAGGTTCCG

NotI
~~~~~
981 AAGGAGTTC AAATGGCGGC CGCTCCACG CAGGGCATCT CCGAAGACCT CTACAGCCCTT TTAGTCGAAA
TTCTCTCAAGG TTTTACCGCG CCGGAGGTGC GTCCCGTAGA GGCTTCTGGA GATGTCGGCA AATCAGCTTT

Sall
~~~~~
1051 TGGCCACTAT CTCCCAAGCT GCCTACGCCG ACCTGTGCAA CATTCGGTCG ACTATTATCA AGGGAGAGAA
ACCGGTGATA GAGGGTTTGA CGGATGCGGC TGGACACGTT GTAAGGCAGC TGATAATAGT TCCCTCTCTT

AccI
~~~~~
BamHI
~~~~~
1121 AATTACAAT TCTCAACTG ACATTAACGG ATGGATCCTC CGCGACGACA GCAGCAAAGA AATAATCACC
TTAAATGTTA AGAGTTTGAC TGTAATTGCC TACCTAGGAG GCGCTGCTGT CGTCTGTTCT TTATTAGTGG

1191 GTCTCCGTG GCACTGGTAG TGATACGAAT CTACAACCTG ATACTAACTA CACCCCTCAC CCTTTTCGACA
CAGAAGGCAC CGTGACCATC ACTATGCTTA GATGTTGAGC TATGATTGAT GTGGGAGTGC GGAAAGCTGT

1261 CCTTACCACA ATGCACCGGT TGTGAAGTAC ACGGTGGATA TTATATTGGA TGGGTCTCCG TCCAGGAGCA
GGGATGGTGT TACGTTGCGA ACACCTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCTCTGGT

```

FIG.\_46C

```

1331 AGTCGAGTCG CTTGTCAAAC AGCAGGTTAG CCAGTATCCG GACTACGCGC TGACCGTGAC CGCCACACCC
    TCAGCTCAGC GAACAGTTTG TCGTCCAATC GGTCAATAGG CTGATGCGCG ACTGGCACTG GCCGGTQMGG

1401 CTCGGCGCCT CCCTGGCGGC ACTCACTGCC GCCCAGCTGT CTGCGACATA CGACAACATC CGCCTGTACA
    GAGCCGCGGA GGGACCGCCG TGAGTGACGG CGGTCGACA GACCTGTAT GCTGTTGTAG GCGGACATGT

                                XhoI
                                ~~~~~
1471 CCTTCGGCGA ACCGCGCAGC GGCAATCAGG CTTTCGCGTC GTACATGAAC GATGCTTCC AAGCCTCGAG
 GGAAGCCGCT TGGCGCGTCG CCGTTAGTCC GGAAGCGCAG CATGTACTTG CTACGGAAGG TTCGGAGCTC

1541 CCCAGATACG ACCCAGTATT TCCGGGTAC TCAATGCCAAC GACGGCATCC CAACCTGCC CCCGGTGGAG
 GGGTCTATGC TCGTCAATA AGGCCCAGTG AGTACGGTTG CTGCCGTAGG GTTTGGACGG GGGCCACCTC

 NcoI
                                ~~~~~
1611 CAGGGGTACG CCCATGGCGG TGTAGAGTAC TGGAGCGTTG ATCCTTACAG CGCCCAAGAAC ACATTGTCT
    GTCCCATGTC GGGTACCGCC ACAFTCATG ACCTCGCAAC TAGGAATGTC GCGGCTCTTG TGTAAACAGA

1681 GCACTGGGGA TGAAGTGCAG TGTGTGAGG CCCAGGGCGG ACAGGGTGTG AATAATGCGC ACACGACTTA
    CGTGACCCCT ACTTCACGTC ACGACACTCC GGGTCCCBCG TGTCCCACAC TTATTACGGG TGTGCTGAAT

                                SphI
                                ~~~~~
 NotI
                                ~~~~~
1751 TTTTGGGATG ACGAGCGGCG CAGCACCTG GCCGCTCGCG GCCCGGAAA CCACTGAAGG ATGAGCTGTA
    AAAACCCCTAC TGCTCGCCGC GTACCTGGAC CGGCCAGCGC CGGCGCCCTT GGTGACTTCC TACTCGACAT

1821 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT CCTGTGCGCG GTCTTTCGAT
    TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA GGACAACGGC CAGAACGCTA

1891 GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA AATAATACA TGTAAATGCAT GACGTTATTT
    CTAATAGTAT ATTAAGACA ACTTAATGCA ATTGCTACAT TATTAAATTGT ACATTACGTA CTGCAATAAA

```

FIG. 46D

```

1961  ATGAGATGGG  TTTTATATGAT  TAGAGTCCCG  CAATTATACA  TTTAATACGC  GATAGAAAAC  AAAATATAGC
      TACTCTACCC  AAAAATACTA  ATCTCAGGGC  GTTAATATGT  AAATTATGCG  CTATCTTTTG  TTTTATATCG
      BssHII
      ~~~~~
 XbaI
      ~~~~~
2031  GCGCAAACTA  GGATAAATTA  TCGCGCGCGG  TGTCACTCTAT  GTTACTAGAT  CGATAAGCTT  CTAGAGCGGC
      CGCGTTTGAT  CCTATTTAAT  AGCGCGCGCC  ACAGTAGATA  CAATGATCTA  GCTATTGAA  GATCTCGCCG
      BssHII
      ~~~~~
 ClaI HindIII
      ~~~~~
2101  CGGTGGAGCT  CCAATTCGCC  CTATAGTGAG  TCGTATTACG  CGCGCTCACT  GCGGTGCTT  TTACAACGTC
      GCCACCTCGA  GGTAAAGCGG  GATATCACTC  AGCATAATGC  GCGCGAGTGA  CCGGCAGCAA  AATGTTGCAG
2171  GTGACTGGGA  AAACCCCTGGC  GTTACCCAAC  TTAATCGCCT  TGCAGCACAT  CCCCTTTTCG  CCAGCTGGCG
      CACTGACCCT  TTTGGGACCG  CAATGGGTTG  AATTAGCGGA  ACGTCGTGTA  GGGGAAAAGC  GGTCTGACCGC
      BssHII
      ~~~~~
2241 TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG
 ATTATCGCTT CTCCGGGGCGT GGCTAGCGGG AAGGTTGTC AACGCGTCGG ACTTACCGCT TACCCCTGCGC
2311 CCTGTAGCG GCGCAATTAAG CGCGCGGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG
 GGGACATCGC CGCGTAATTC GCGCGGCCCA CACCACCAAT GCGCGTCGCA CTGGCGATGT GAACGGTTCG
2381 CCTAGCGCC CGCTCCTTTTC GCTTTCTTTC CTTCTCTTCT CGCCACGFTC GCCGGCTTTC CCCGTCGAAGC
 GGGATCGCGG GCGAGGAAAG CGAAGAAGG GAAGGAAGA GCGGTGCAAG CGGCCGAAAG GGGCAGTTTCG
2451 TCTAAATCGG GGGCTCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGAT
 AGATTTAGCC CCCGAGGGAA ATCCCAAGGC TAAATCAGGA AATGCCGTGG AGCTGGGGTT TTTTGAACTA
2521 TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTCG CCCTTTGACG TTGGAGTCCA
 ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAGC GGGAAACTGC AACCTCAGGT

```

FIG. 46E

2591 CGTCTTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCCT ATCTCGGTCT ATTCTTTTGA  
 GCAAGAAATT ATCACCCTGAG AACAAGGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACCT  
 2661 TTTATAAAGGG ATTTTGCCGA TTTCGGCCTA TTGTTAAAA AATGAGCTGA TTTAACAAAA ATTAAACGCG  
 AAATATTCCC TAAACGGCT AAAGCCGGAT AACCAATTTT TTAATCGACT AAATTTGTTT TAAATTTGCGC  
 2731 AATTTTAACA AAATATTAAAC GCTTACAATT TAGGTGGCAC TTTTTCGGGA AATGTGCGCG GAACCCCTAT  
 TTAATAATTGT TTTATAATTG CGAATGTTAA ATCCACCGTG AAAAGCCCTT TTACACGCGC CTTGGGGATA  
 2801 TTGTTTATTT TTCTAATAAC ATTCAAAATAT GTATCCGCTC ATGAGACAAT AACCTTGATA AATGCTTCAA  
 AACAAATAAA AAGATTATG TAAGTTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAGTT  
 2871 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTGCGCCT TATTCCTTTT TTTGCGGCAT  
 ATTATTAATT TTTCTTCTC ATACTCATAA GTTGTAAAGG CACAGCGGA ATAAAGGAAA AAACGCCGTA  
 2941 TTTGCCCTCC TGTTTTNGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC  
 AAACGGAAGG ACRAAAACGA GTGGGTCTTT GCGACCACIT TCATTTTCTA CGACTTCTAG TCAACCCACG  
 3011 ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTTCGCC CGAAGAACGT  
 TGCTCACCCA ATGTAGCTTG ACCTAGAGTT GTCGCCATTG TAGGAACCTT CAAAAGCGGG GCTTCTTGCA  
 3081 TTTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG  
 AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCATAACTG CGGCCCGTTC  
 3151 AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACIT GGTGAGTAC TCACCAGTCA CAGAAAAAGCA  
 TCGTTGAGCC AGCGGCGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTTCGT  
 3221 TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC  
 AGAATGCCTA CCGTACTGTC ATTCTCTTAA TACGTCACGA CGGTATTGGT ACTCACTATT GTGACGCGCGG  
 3291 AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCCTTTTTT GCACAACATG GGGGATCATG  
 TTGAATGAAG ACTGTTGCTA GCCTCCTGGC TTCTCTCGATT GCGGAAAAAA CGTGTGTGAC CCCCTAGTAC  
 3361 TAACTCGCCT TGAATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCCAGAT  
 ATTGAGCGGA ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTGTGCTA

**FIG. 46F**



3431 GCCTGTAGCA ATGCAACAA CGTTGCGCAA ACTATTAACT GGCAACTAC TTACTCTAGC TTCCCGGCAA  
CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT

3501 CAATTAATAG ACTGGATGA GCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT  
GTTAATTATC TGACCTACCT CCGCTTATTT CAACGTCTTG GTGAAGACGC GAGCCGGGAA GGCCGACCGA

3571 GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGGCCAGA  
CCAAATAACG ACTAATTAGA CCTGGGCCAC TCGCACCCAG AGCGCATAG TAACGTCTGT ACCCCGCTCT

3641 TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA  
ACCATTCGGG AGGCAATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCTTTATCT

3711 CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT AACTGTGAGA CCAAGTTTAC TCATATATAC  
GTCTAGCGAC TCTATCCACG GAGTGACTAA TTGTAACCA TTGACAGTCT GGTTCAAATG AGTATATATG

3781 TTTAGATTGA TTTAAACTT CATTTTAAAT TTTAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT  
AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCTA GATCCACTTC TAGGAAAAAC TATTAGAGTA

3851 GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
CTGGTTTTAG GGAATTGCAC TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA

3921 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAA ACCACCGCTA CCAGCGGTGG  
AGAACTCTAG GAAAAAAGA CCGCATTAG ACGACGAACG TTTGTTTTTT TGGTGGCGAT GGTCCGCCACC

3991 TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCGGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC  
AAACAAACGG CCTAGTTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG

4061 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAATCT CTGTAGCACC GCCTACATAC  
TTTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCTTGA GACATCGTGG CGGATGTATG

4131 CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT  
GAGCGAGACG ATTAGGACAA TGGTCAACCGA CGACGGTCAAC CGCTATTCTAG CACAGAATGG CCCAACCTGA

4201 CAAGACGATA GTTACCGGAT AAGCGCAGC GGTCTGGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT  
GTTCTGCTAT CAATGGCCTA TTCCGCGTCTG CCAGCCCGAC TTGCCCCCA AGCACGTGTG TCGGGTCTGAA

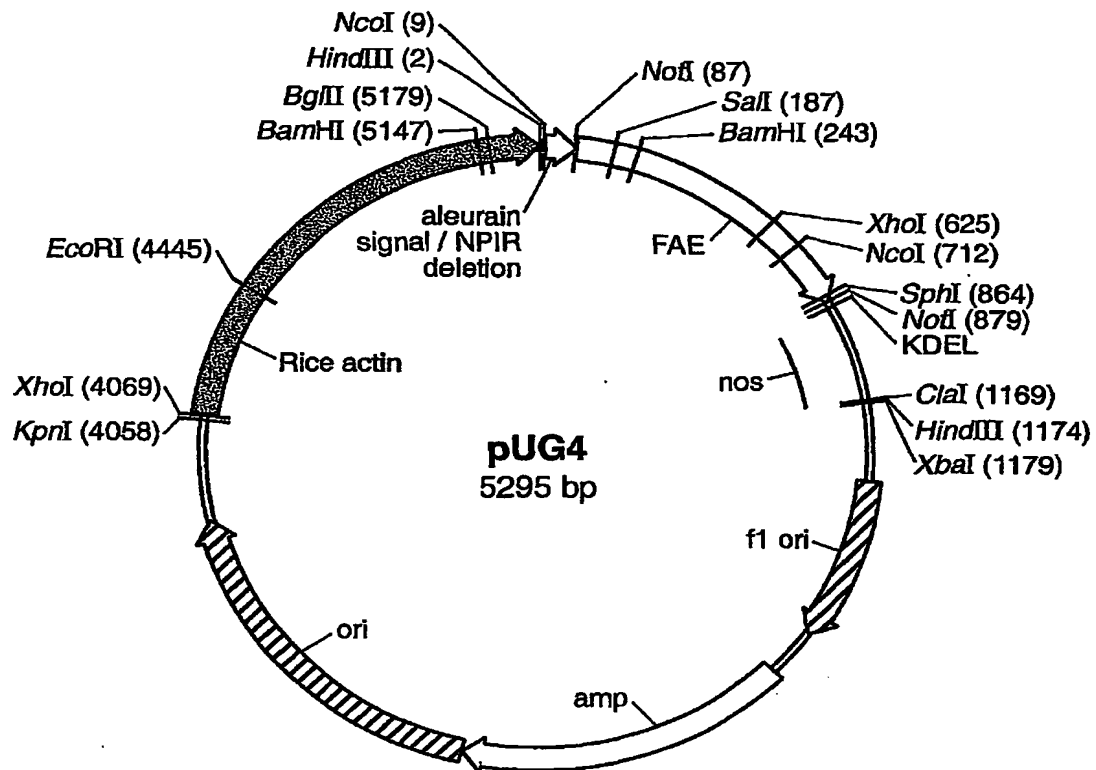
**FIG.-46G**

4271 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA  
 CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGAATGTGCA CTCGATATC TTTCGGCGTG CGAAGGGCTT  
 4341 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG  
 CCCTCTTTCC GCCTGTCCAT AGGCCATTG AGCCCCAGC CTTGTCTCTT CCGTGTCTCC CTCGAAGGTC  
 4411 GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG  
 CCCCTTTGCG GACCATAGAA ATATCAGGAC AGCCCAAGC GGTGGAGACT GAACTCGCAG CTAAAAACAC  
 4481 ATGCTCGTCA GGGGGCGGA GCCTATGGAA AAACGCCAGC AACGGGCGCT TTTTACGGTT CCTGGCCCTT  
 TACGAGCAGT CCCCCCGCCT CGGATACCTT TTTGCGGTG TTGCGCCGGA AAATGCCAA GGACCCGGAA  
 4551 TGTGCGCCTT TTGCTCACAT GTTCTTTCTT GCGTTATCCC CTGATTTCTGT GGATAACCGT ATTACCGCCT  
 ACGACCGGAA AACGAGTGTA CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA  
 4621 TTGAGTGAGC TGATACCGCT CGCCCGAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA  
 AACTCACTCG ACTATGGCGA GCGGCGTCGG CTGTCTGGCT CGCGTCGCTC AGTCACTCGC TCCTTTCGCC  
 4691 AGAGCGCCCA ATACGCAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCACTG GCACGACAGG  
 TCTCGCGGGT TATGCGTTTG GCGGAGAGGG GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGTGTCTC  
 4761 TTTCCCGACT GGAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC  
 AAAGGCTGA CCTTTCGCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA ATCCGTGGGG  
 4831 AGGCTTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGTG AATTGTGAGC GGATAACAAT TTCACACAGG  
 TCCGAAATGT GAAATACGAA GCGCGAGCAT ACAACACACC TTAACACTCG CCTATTGTTA AGTGTGTCTC

BssHII  
 ~~~~~  
 EcoRI

4901 AAACAGCTAT GACCATGATT ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGG  
 TTTGTGCGATA CTGGTACTAA TCGGTTTCGC GCGTTAATTG GGAATGATTT CCCTTGTCTT CGACC

**FIG. 46H**

**FIG. 47A**

```

 NcoI
      ~~~~~
HindIII
~~~~~
 M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGCGC TCGCCGTGCT GGCACGGCC GCGTCGCCG

      ~~~~~
      NotI
      ~~~~~
 . A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCTCCTC CCGCGCGCC GCCTCCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAT

      ~~~~~
      Sall
      ~~~~~
 . A T I S Q A A Y A D L C N I P S T I I K G E K
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA GGGAGGAAA

      ~~~~~
      BamHI
      ~~~~~
 I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTACAAAT CTCAACTGA CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG
 . F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAATAC ACCCTCACGC CTTTCGACAC
 . L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA
 V E S L V K Q Q V S Q Y P D Y A L T V T G H K L
421 GTCGAGTCGC TTGTCAACA GCAGGTTAGC CAGTATCCGG ACTACGGCT GACCGTGACC GGCCACKCCC
 . G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGCGGCCCTC CCTGGCGCA CTCACCTGCC CCCAGCTGTC TCGGACATAC GACACATCC GCCTGTATAC

      ~~~~~
      XhoI
      ~~~~~
 . F G E P R S G N Q A F A S Y M N D A F Q A S S
561 CTTGGGCGAA CCGCGCAGCG GCAATCAGGC CTTGCGCTCG TACATGAACG ATGCTTCCA AGCCTCGAGC
 P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAAGC ACGCATCCC AAACCTGCC CCGGTGGAGC

      ~~~~~
      NcoI
      ~~~~~

```

FIG.\_47B

. G Y A H G G V E Y W S V D P Y S A Q N T F V C .  
 701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACCA CATTTGCTCG  
 . T G D E V Q C C E A Q G G Q G V N A H T T Y  
 771 CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT

NotI

SphI

F G M T S G A C T W P V A A A E P L K D E L \*  
 841 TTGGGATGA CGAGCGGCG ATGCACCTGG CCGTCGCGG CCGCGGAACC ACTGAAGGAT GAGCTGTAAA  
 911 GAAGCAGATC GTTCAAAACAT TTGGCAATAA AGTTTCTTAA GATTGAATCC TGTTCGCCGT CTTGCGATGA  
 981 TTATCATATA ATTCTGTGTG AATTACGTTA AGCATGTAAT AATTAACATG TAATGCAATG CGTTATTAT  
 1051 GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAACAA AATATAGCGC

HindIII

ClaI

XbaI

1121 GCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT TACTAGATCG ATAGCTTCT AGAGCGGCGG  
 1191 GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGG GTCTACACTGG CCGTCGTTT ACAACGTCGT  
 1261 GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG CAGCACATCC CCGTTTCGCC AGCTGGCGTA  
 1331 ATAGCGAAGA GCGCCGCACC GATCGCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGACCGCGCC  
 1401 CTGTAGCGGC GCATTAAAGC CGCGGGGTGT GGTGGTTACG CCGAGCGTGA CCGCTACACT TCGCAGCGCC  
 1471 CTAGCGCCCG CTCCTTTCGC TTTCTTCCTT TCCCTTCTCG CCACGTTCCG CCGCTTTCCT CGTCAAGCTC  
 1541 TAAATCGGGG GCTCCCTTTA GGGTTCCGAT TTAGTGTCTT ACGGCACCTC GACCCCAAA AACTTGATTA  
 1611 GGGTGTAGGT TCACGTAGTG GGCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG  
 1681 TTTCTTAAAT GTGGACTCTT TCGGCTATT TCGTAAATAA TGAGCTGATT TACACAAAAT TTAACGCGAA  
 1751 TATAAGGGAT TTTGCCGATT TTACAAATTA GGTGGCACTT TTCGGGGAAA TGTGCGCGGA ACCCTATT  
 1821 TTTTAAACAA ATATTAAACG TTACAAATTA ATCCGCTCAT GAGACAAATA CCTGATATA TGCCTCAATA  
 1891 GTTTATTATT CTAATACAT TCAATATGT TGAGTATCA ACATTTCCTG GTCCGCTTA TCCCTTTT TCGGGCATTT  
 1961 ATATTGAAA AGGAAGAGTA TGAGTATCA CCCAGAAACG CTGGTGAAG CCGTGAAGT CTTTGAGAGT AAGAACGTTT  
 2031 TGCCTTCCTG TTTTTCCTCA CATCGMACTG GATCTCAACA CCGTGAAGT CTTTGAGAGT TTTGCGCCCG AAGAACGTTT  
 2101 GAGTGGGTTA AGCACTTTTA AAGTCTGCT ATGTGGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG  
 2171 TCCAATGATG GCGCATACA CTATTCTCAT AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC  
 2241 CAACCTCGGC CATGACAGTA AGAGATTAT GAGCTAAC CCATTTCCTG CATAACATG AGTGAATAACA CTGCGGCCAA  
 2311 TTACGGATGG ACAACGATCG GAGGACCGAA CGAGCTAAC GCTTTTTCG ACAACATGGG GATCATGTA  
 2381 CTACTTCTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAACGGA CAGCTGTGAC ACCACGATGC  
 2451 ACTCGCCTTG GGCACAAACG TTGCGCAAC TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA  
 2521 CTGTAGCAAT

FIG.-47C

|       |             |             |            |             |             |             |             |
|-------|-------------|-------------|------------|-------------|-------------|-------------|-------------|
| 2591  | ATTAATAGAC  | TGGATGGAGG  | CGGATAAAGT | TGCAGGACCA  | CTTCTGCGCT  | CGGCCCTTCC  | GGCTGGCTGG  |
| 2661  | TTTATTGCTG  | ATAAATCTGG  | AGCGGTGAG  | CGTGGGTCTC  | GCGGTATCAT  | TGCAGCACTG  | GGGCCAGATG  |
| 2731  | GTAAGCCCTC  | CCGTATCGTA  | GTTATCTACA | CGACGGGGAG  | TCAGGCAACT  | ATGGATGAAC  | GAAATAGACA  |
| 2801  | GATCGCTGAG  | ATAGGTGCCCT | CACGTGATTA | GCAATGGTAA  | CTGTGAGACC  | AGTTTACTC   | ATATATACCT  |
| 2871  | TAGATTGATT  | TAAACCTTCA  | TTTTTAATTT | AAAAGGATCT  | AGGTGAAGAT  | CCTTTTGTAT  | AATCTCATGA  |
| 2941  | CCAAAATCCC  | TTAACGCTGAG | TTTTTCTGTC | ACTGAGCGTC  | AGACCCCGTA  | GAAAAGATCA  | AAGGATCTTC  |
| 3011  | TTGAGATCCT  | TTTTTCTGTC  | GCGTAATCTG | CTGCTTGCAA  | ACAAAAAATC  | CACCGGTACC  | AGCGGTGGTT  |
| 3081  | TGTTTGCCCG  | ATCAAGAGCT  | ACCAACTCTT | TTTCCGGAAG  | TAACGTGGCT  | GAGCAGAGCG  | CAGATACCAA  |
| 3151  | ATACTGTCTT  | TCTAGTGTAG  | CCGTAGTTAG | GCCACCACCT  | CAAGAATCTT  | GTAGCACCGC  | CTACATACCT  |
| 3221  | CGCTCTGCTA  | ATCCTGTTAC  | CAGTGGCTGC | TGCCAGTGGC  | GATAAGTCTG  | GTCTTACCGG  | GTTGGACTCA  |
| 3291  | AGACGATAGT  | TACCGGATTA  | GGCGCAGCGG | TCGGGCTGAA  | CGGGGGGTTC  | GTGCACACAG  | CCGAGCTTGG  |
| 3361  | AGCGAACGAC  | CTACACCCGAA | CTGAGATACC | TACAGCGTGA  | GCTATGAGAA  | AGCGCCACGC  | TTCCCGAAGG  |
| 3431  | GAGAAAGGCG  | GACAGGTATC  | CGGTAAGCGG | CAGGTCGGA   | ACAGGAGAGC  | GCACGAGGGA  | GCTTCCAGGG  |
| 3501  | GGAAACGCCCT | GGTATCTTTA  | TAGTCTGTCT | GGTTCGCG    | ACCTCTGACT  | TGAGCGTCTGA | TTTTTGTGAT  |
| 3571  | GCTCGTCAGG  | GGGCGGAGC   | CTATGGAAAA | ACGCCAGCAA  | CGCGGCCCTTT | TTACGGTTCC  | TGGCCCTTTT  |
| 3641  | CTGGCCTTTT  | GCTCACATGT  | TCTTCTCTGC | GTTATCCCTT  | GATCTGTGG   | ATAACCGTAT  | TACCGCCCTTT |
| 3711  | GAGTGAGCTG  | ATACCGCTCG  | CCGACGCGCA | ACGACCGAGC  | GCAGCGAGTC  | AGTGAAGGAG  | GAAAGCGGAG  |
| 3781  | AGCGCCCAAT  | ACGCCAAACG  | CCTCTCCCGG | CGCGTTGGCC  | GATTCATTA   | TGCAGCTGGC  | ACGACAGGTT  |
| 3851  | TCCCGACTGG  | AAAGCGGGCA  | GTGAGCGCAA | CGCAATTAAT  | GTGAGTTAGC  | TCACTCATTA  | GGCACCCCCAG |
| 3921  | GCTTTACACT  | TTATGCTTCC  | GGCTCGTATG | TTGTGTGGA   | TTGTGAGCGG  | ATAACAATTT  | CACACAGGAA  |
| KpnI  |             |             |            |             |             |             |             |
| 3991  | ACAGCTATGA  | CCATGATTAC  | GCCAAGCGCG | CAATTAAACC  | TCACTAAAGG  | GAAACAAAGC  | TGGGTACCCG  |
| XhoI  |             |             |            |             |             |             |             |
| 4061  | GCCCCCCTC   | GAGGTCAATC  | ATATGCTTGA | GAAGAGAGTC  | GGGATAGTCC  | AAAATAAATC  | AAAGGTAGA   |
| 4131  | TTACCTGGTC  | AAAAGTGAAA  | ACATCAGTTA | AAAGGTGGTA  | TAACTAAAAT  | ATCGGTAATA  | AAAGGTGGCC  |
| 4201  | CAAAGTGAAA  | TTTACTCTTT  | TCTACTATTA | TAAAAATTGA  | GGATGTTTGG  | TCGGTACTTT  | GATACGTCTAT |
| 4271  | TTTTTGATGA  | ATTGGTTTTT  | AAGTTTATTC | GGGATTTGGA  | AATGCATATC  | TGTATTGTAG  | TCGGTTTTTA  |
| 4341  | AGTTCGTTGC  | TTTTGTAAAT  | ACAGAGGGAT | TTGTATAAGA  | AATATCTTTA  | AAAAACCCAT  | ATGCTAATTT  |
| EcoRI |             |             |            |             |             |             |             |
| 4411  | GACATAATTT  | TTGAGAAAAA  | TATATATTCA | GGCGAATTC   | ACAATGAACA  | ATAATAAGAT  | TAAAAATAGCT |
| 4481  | TGCCCCCGTT  | GCAGCGATGG  | GTATTTTTTC | TAGTAAAAAT  | AAAGATAAAC  | TTAGACTCAA  | AACATTTACA  |
| 4551  | AAACAAACCC  | CTAAAGTCTT  | AAAGCCCAAA | GTGCTATGCA  | CGATCCATAG  | CAAGCCCAAG  | CCAACCCCAAC |
| 4621  | CCAAACCCAC  | CCAGCCCACT  | GCAGCCCACT | GGCAATATAGT | CTCCACCCCC  | GGCACTATCA  | CCGTGAGTTG  |

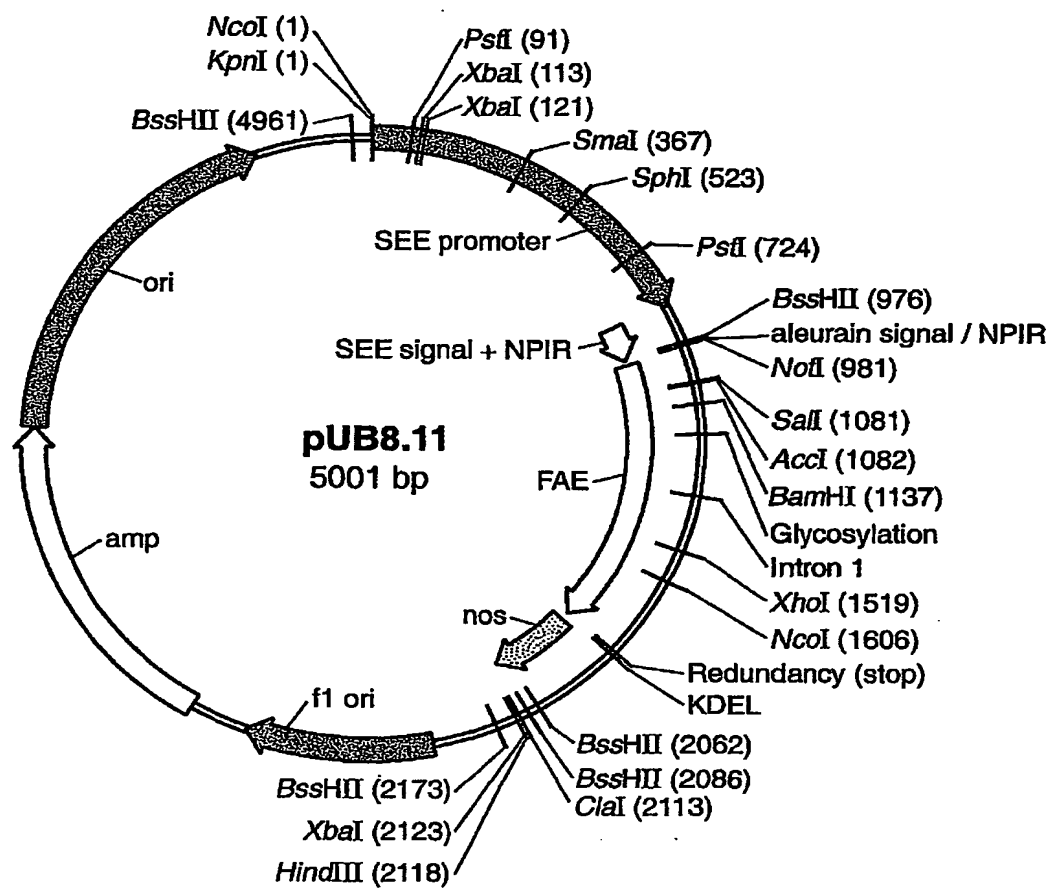
FIG. 47D

```

4691 TCCGCACCAC CGCACGTCTC GCAGCCAAAA AAAAAAAG AAGAAAAAAA AAAAAAAGA AAAACAGCAG
4761 GTGGGTCCGG GTCTGTGGGG CCGGAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC
4831 CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCACAACC
4901 TACCACCACC ACCACCACCA CTTCTTCCCC CTTGCTGCC GGAACGACGAG CTCTTCCCCC CTCCCCCTCC
4971 GCCGCCGCCG GTAACCAACC CGCCCTCTC CTTCTTCTT CTTCTTCTT CTTCTTCTT TTTTTCGTCT CGGTCTCGAT
5041 CTTTGGCCCT GGTAGTTTGG GTGGGCGAGA GCGGCTTCTG BsmHI CGCCCAAGTC GTGCGCGGG AGGGGCGGGA
 BglII
5111 TCTCGCGGCT GGCGTCTCCG GGCGTGAATC GGCCCGGATC CTCGCGGGGA ATGGGGCTCT CGGATGTAGA
 BglII
5181 TCTTCTTCTT TTTCTTCTTT TGTGGTAGAA TTGGAATCCC TCAGCATGCT TCATCGGTAG TTTTCTCTTT
5251 CATGATTTGT GACAAATGCA GCTTCGTGCG GAGCTTTTGT GTAGC

```

FIG.-47E

**FIG. 48A**



NcoI  
 ~~~~~  
 KpnI  
 ~

1 CATTGGGCCAG GTATATATAT GGGATATCTC AAGCAATATA TCGAATATATC ACCATTGGCT ACAATATCTG  
 GTACCCGGTC CATATTAATA CCTATATAG CCCTATATAG TTCGTTTATT AGCTTATATAG TGGTAACCGA TGTATATAGAC

PstI  
 ~~~~~

71 AGTCCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAATC CTAGATAGCA CAGCCACAGC  
 TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAAAC TAGATCTTGA GATCTATCGT GTCGGTGTCTG

XbaI XbaI  
 ~~~~~

141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG TCATCACAAAC CTCTGCCTCG AGAAGGATG CTCCTGACGT  
 TGGATGTCTT CACGCTGTGA ACACCTGACA TCGTCAACCA TCACGCGCTC CCAACAAAT ATCGTCCCCC

211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAT ATCGTCCCCC  
 ACGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTCCGCGAG GTTTGTTTTA TAGCAGGGGG

281 ATGCTTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTGCT GAATCTCGCT TCCACTGGCC  
 TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAAACAGA CTTAGAGCGA AGGTGACCGG

SmaI  
 ~~~~~

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGGA  
 TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCCT

421 TGGAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCCAAT  
 ACCTTTCTTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA

SphI  
 ~~~~~

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TCGTTTCGGG TTTTGTGTTGG CTTAATTGAC TTTATTTTGG  
 AGTCTAAGGG GGTATCTCTT TTCTATCTCT ACGAAAGCCC AAAACAAACC GAATTAAC TG AATAAAAAAC

561 TTGAGATTGA ATGCTGATTT GTTGTGTAAT ATGCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG  
 AACCTCAACT TAGGACTAAA CAACACATTT TACGGGTGG TAGACTTATA GCTCTGCCTA TTATCCGACC

FIG. 48B

```

631 CTAATAAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

 PstI
      ~~~~~
701  TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
    AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCTGTCACC ACTTCGCACC GCTACTCTAC CCATATTTTG

771  CCCCAGCACC GGGACGCGAG CTCCTGCGCTA CCAGTACCAT CTCGCCCTGC TCCCCCTGCC GGACGACCCA
    GGGCCCGTGG CCTGCGCTC GAGGCGGAT GGTCAAGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841  GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
    CATTTTATGA CAACGGGTGA CGGCCCGCTC TACCGGGTGC CGCGTAGGA GAAGAACCSC GAGCGGCAGA

      BssHII
      ~~~~~
 NotI
      ~~~
911  TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
    ACCGGTGGCG CGGCCACCGG CGCGGTAGNA ACCGCCTGAG GTTGGGCTAG GCCGGGCTGT GGCCTGCGCG

      NotI
      ~~~~~
981 GGCCGCCCTCC ACGCAGGGCA TCTCCGAGA CCTCTACAGC CGTTAGTCTG AATGGCCAC TATCTCCCAA
 CCGCGGAGG TCGTCCCGT AGAGGCTTCT GGAGATGTCG GCAANTCAGC TTACCCGGTG ATAGAGGGTT

 Sali
      ~~~~~
      AccI
      ~~~~~
1051 GCTGCCCTACG CCGACCTGTG CAACATTCCG TCGACTATTA TCAAGGGAGA GAAATTTAC AATTCCTCAA
 CGACGGATGC GGCTGGACAC GTTGTAAAGC AGCTGATAAT AGTTCCCTCT CTTTAAATG TTAAGAGTTT

```

**FIG. 48C**

BamHI  
 ~~~~~  
 1121 CTGACATTAA CGGATGGATC CTCGGGACG ACAGCAGCAA AGAATAATC ACCGCTTCC GTGGCACTGG  
 GACTGTAAAT GCCTACCTAG GAGCGCTGC TGTCTCGTT TCTTTATTAG TGGCAGAAG CACCGTGACC  
  
 1191 TAGTGATACG AATCTACAAC TCGATACTAA CTACACCTTC ACGCCTTTC ACACCTTACC ACAATGCAAC  
 ATCACTATGC TTAGATGTTG AGCTATGAT GATGTGGAG TCGGGAAGC TGTGGATGG TGTACGTTG  
  
 1261 GGTGTGAAG TACACGGTGG ATATTATATT GGATGGTCT CCGTCCAGGA CCAAGTCCG TCGCTTGTCA  
 CCAACACTTC ATGTGCCACC TATAATATAA CCTACCCAGA GGCAGGTCCT GGTTCAGTC AGCGAACAGT  
  
 1331 AACAGCAGGT TAGCCAGTAT CCGGACTACG CGCTGACCGT GACCGGCCAC KCCCTCGGC CCTCCCTGGC  
 TTGTCTGTTCA ATCGGTCTATA GGCTGTATG CCGACTGGCA CTGGCCGCTG MGGGAGCCGC GGAGGGACCG  
  
 1401 GGCACCTACT GCCGCCCAGC TGTCTGGAC ATACGACAAC ATCCGCTGT ACACCTTGG CGAACCGCGC  
 CCGTGAATGA CCGCGGGTGC ACAGACGCTG TATGCTGTTG TAGCGGACA TGTGAAGCC GCTTGGCGCG  
  
 XbaI  
 ~~~~~  
 1471 AGCGGCAATC AGGCTTTCG CTCGTACATG AACGATGCTT TCCAGCCTC GAGCCAGAT ACGACGCAGT  
 TCGCCGTTAG TCCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTCCGAG CTCGGGTCTA TGCTGCGTCA  
  
 NcoI  
 ~~~~~  
 1541 ATTTCCGGGT CACTCATGCC AACGACGGCA TCCCAAACCT GCCCCCGGTG GAGCAGGGGT ACGCCCATGG  
 TAAAGGCCCA GTGAGTACGG TTGCTGCCGT AGGTTTGA CCGGGGCCAC CTCGTCCCA TCGGGGTACC  
  
 1611 CGGTGTAGAG TACTGGAGCG TTGATCTTA CAGCGCCAG AACACATTG TCTGCACTGG GATGAAGTG  
 GCCACATCTC ATGACCTCGC AACTAGGAAT GTCCGGGGTC TTGTGTAAC AGACGTGACC CCTACTTTCAC  
  
 1681 CAGTGTGTG AGGCCCAGGG CGGACAGGCT GTGAATAATG CGCACACGAC TTATTTTGGG ATGACGAGCG  
 GTCACGACAC TCCGGGTCCC GCCTGTCCA CACTTATTAC GCCTGTGCTG AATAAACCC TACTGCTCGC  
  
 1751 GAGCCTGTAC ATGGTGATCA GTCATTTTAC CCTCCCCGAG TGTACCAGGA AAGATGGATG TCCTGGAGAG  
 CTCGGACATG TACCACATAG CAGTAAAGTC GGAGGGCTC ACATGGTCTT TTCTACCTAC AGGACCTCTC

FIG. 48D

```

1821 GGGGCCGCGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTTCAA ACATTGGCA ATAAAGTTTC
CCCCGGCGCA TTGGTGACTT CCTACTCGAC ATTCTTTCGT CTAGCAAGTT TGTAAACCGT TATTTCAAAG

1891 TTAAGATTGA ATCCTGTTGC CGGTCTTGCG ATGATTATCA TATAATTTC TTTGAATTAC GTTAAGCATG
AATCTAACT TAGGACAACG GCCAGAACGC TACTAATAGT ATATTAAAG CAACTTAATG CAATTCGTAC

1961 TAATAATTAA CATGTAATGC ATGACCTTAT TTATGAGATG GGTTTTATG ATTAGAGTCC CGCAATTATA
ATTATTAAAT GTACATTACG TACTGCAATA AATACTCTAC CCAAAAATAC TAATCTCAGG GCGTTAATAT

 BssHII
                                     ~~~~~
2031 CATTTAATAC GCGATAGAAA ACAAAATATA GCGCGCAAC TAGGATAAAT TATCGCGCGC GGTGTCACTT
GTAAATTATG CGCTATCTTT TGTTTTATAT CGCGCGTTTG ATCCTATTTA ATAGCGCGCG CCACAGTAGA

                                     BssHII
                                     ~~~~~
 xbaI
                                     ~~~~~
                                     ClaI HindIII
                                     ~~~~~
2101 ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGGTGGAG CTCCAATTCC CCTATATAGT AGTCGTATTA
TACAATGATC TAGCTATTTC AAGATCTCGC CGGCCACCTC GAGGTTAAGC GGGATATCAC TCAGCATAAAT

 BssHII
                                     ~~~~~
2171 CGCGCGCTCA CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCCTG GCGTTACCCA ACTTAATCGC
GCGCGCGAGT GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTGTGGGAC CGCAATGGGT TGAATTAGCG

2241 CTTGCAGCAC ATCCCCCTTT CGCCAGTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC
GAACGTCGTG TAGGGGAAA GCGGTCGACC GCATTATCGC TTCTCCGGGC GTGGCTAGCG GGAAGGGTTG

2311 AGTTGCGCAG CCTGAATGGC GAATGGGAG GCGCCTGTAG CGGCGCATTA AGCGCGCGCG GTGTGGTGGT
TCAACGCGTC GGACTTACCG CTTACCCCTGC GCGGACATC GCCGCGTAAT TCGCGCGGCC CACACCACCA

2381 TAGCGGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCCTTCCCTT
ATGCGCGTCG CACTGGCGAT GTGAACGTC GCGGATCGC GGGCGAGGAA AGCGAAGAA GGAAGGAAA

```

**FIG. 48E**

2451 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG  
GAGCGGTGCA AGCGGCCGAA AGGGCGAGTT CGAGATTAG CCCCGAGGG AATCCCAAG GCTAAATCAC

2521 CTTTACGGCA CCTCGACCCC AAAAAACTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA  
GAAATGCCGT GGAGCTGGGG TTTTITGAACT TAATCCCACT ACCAAGTGCA TCACCCGGTA GCGGGACTAT

2591 GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CAGCTCTTTT AATAGTGGAC TCTTGTTCCTA AACTGGAACA  
CTGCCAAAA GCGGGAACCT GCAACCTCAG GTGCAAGAAA TTATCACCTG AGAACCAAGT TTGACCTTGT

2661 ACACTC AACCTCTATCTCGGT CTATCTCTTTT GATTATTAAG GGATTTTGCC GATTCGGCC TATTGGTTAA  
TGTGAGTTGG GATAGAGCCA GATAAGAAA CTAAATATTC CCTAAACCGG CTAAGCCCGG ATAAACAATT

2731 AAAATGAGCT GATTAAACA AAATTTAACG CGAATTTTAA CAAATATTA ACGCTTACAA TTTAGGTGGC  
TTTTACTCGA CTAAATTGTT TTTAAATGTC GCTTAAAAAT GTTTTATAAT TCGGAATGTT AAATCCACCG

2801 ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAT ACATTCAAAT ATGTATCCGC  
TGAAAAAGCCC CTTTACACGC GCCTTGGGGA TAAACAATA AAAAGATTIA TGTAAGTTTA TACATAGGCG

2871 TCATGAGACA ATAAACCTGA TAATGCTTC AATAATATG AAAAAAGGAG AGTATGAGTA TTCAACATTT  
AGTACTCTGT TATTGGGACT ATTTACGAAG TTATATAAC TTTTCTCTTC TCATACTCAT AAGTTGTAAA

2941 CCGTGTGCCC CTTATTCCCT TTTTTCGGC ATTTTGCCCTT CCTGTTTTTG CTCACCCAGA AACGCTGGTG  
GGCACAGCGG GAATAAGGGA AAAACGCGG TAAAAAGGA GAGTGGGTCT TTGGGACCAC

3011 AAAGTAAAA AGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA  
TTTCAATTTT TACGACTTCT AGTCAACCCA CGTGTCTCACC CAATGTAGCT TGACCTAGAG TTGTGCCCCAT

3081 AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAAGTTC TGCTATGTGG  
TCTAGGAACCT CTCAAAAGCG GGGCTTCTTG CAAAAGTTA CTACTCGTGA AATTTCAAG AGGATACACC

3151 CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTCCCGCA TACACTATTC TCAGAATGAC  
GCGCCATAAT AGGGCATAAC TCGCGCCCGT TCTCGTTGAG CCAGCGGCGT ATGTGATAAG AGTCTTACTG

3221 TTGGTTGAGT ACTCACCAGT CACAGAAAA CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG  
AACCAACTCA TGAGTGGTCA GTGTCTTTTC GTAGAATGCC TACCGTACTG TCATTTCTCTT AATACGTCAC

**FIG.\_48F**

3291 CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC CGAAGGAGCT  
 GACGGTATTG GTACTCACTA TTGTGACGCC GTTGAATGA AGACTGTTCG TAGCCTCCTG GCTTCCTCGA  
 3361 AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACCTCG CTTGATCGTT GGGAAACCGGA GCTGAATGAA  
 TTGGCGAATA AACGTGTGT ACCCCCTAGT ACAATGAGCG GAACTAGCAA CCTTGGCCT CGACTTACTT  
 3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCTCTGAG CAATGGCAAC AACGTTGCGC AAACATATTAA  
 CGGTATGGTT TGCTGCTCGC ACTGTGTGTC TACGGACATC GTTACCGTTG TTGCAACGCG TTGTGATAATT  
 3501 CTGGCGAACT ACTTACTCTA GCTTCCCAGC AACAAATTAAT AGACTGAGT GAGGCGGATA AAGTTGCAGG  
 GACCGCTTGA TGAATGAGAT CGAAGGCCG TTGTTAATTA TCTGACCTAC CTCGCGCCTAT TTCAACGTC  
 3571 ACCACTTCTG CGCTCGGCC TCCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG  
 TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAATAAA CGACTATTTA GACCTCGGCC ACTCGCACCC  
 3641 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG  
 AGAGCGCCAT AGTAACGTG TGACCCCGGT CTACCATTCG GGAGGGCATA GCATCAATAG ATGTGCTGCC  
 3711 GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGTCGC TGAGATAGGT GCCTCAGTGA TTAAGCATTG  
 CCTCAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT AATTCGTAAAC  
 3781 GTAACTGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTTCATTTTAA ATTTAAAGA  
 CATTGACAGT CTGGTTCAAA TGAGTATATA TGAATCTAA CTAAATTTTG AAGTAAAAAT TAAATTTTCC  
 3851 ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG TTCCACTGAG  
 TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGTGTTT AGGGAATGC ACTCAAAAGC AAGGTGACTC  
 3921 CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT  
 GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGAAAAAAA GACGCGCATT AGACGACGAA  
 3991 GCAAACAAAA AAACACCCGC TACCAGCGGT GGTTCGTTG CCGGATCAAG AGCTACCAAC TCTTTTTCGG  
 CGTTTGTGTT TTGTTGGGCG ATGTCGCCA CCACAACAAC GGCCTAGTTC TCGATGGTTG AGAAAAAGC  
 4061 AAGGTAACTG GCTTCAGCAG AGCGCAGATA CCAAACTAG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC  
 TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTTATGAC AGGAAGATCA CATCGGCATC AATCCGGTGG

**FIG.\_48G**

4131 ACTTCAAGAA CTCGTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG  
TGAAGTCTTT GAGACATCGT GCGGATGTA TGGAGGAGA CGATTAGGAC AATGGTCACC GACGACGGTC

4201 TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCGAAGACGA TAGTTACCGG ATAAGGCCCA GCGGTCTGGC  
ACCGCTATTC AGCACAGAA AGCCCAACCT GAGTCTGCT ATCAATGGCC TATTCGCGT TATTCGCGC

4271 TGAACGGGGG GTTCGTGCAC ACAGCCACG TGTGCGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTG  
ACTTGCCCCC CAAGCACGTG TGTGCGGTG TGTGCGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTG

4341 GTGAGCTATG AGAAGCGCC ACGCTTCCG AAGGGAGAA GCGGACAGG TATCCGGTAA GCGGCAGGGT  
CACTCGATAC TCTTTCGCG TCGGAAGGCG TTCCCTCTTT CCGCTGTCC ATAGGCCATT GCGCTGCCA

4411 CGGAACAGGA GAGCGCACG GGGAGCTTCC AGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGTTT  
GCCTTGTCTT CTGCGGTGCT CCTCGAAGG TCCCCCTTG CCGACCATAG AATATCAGG ACAGCCCAA

4481 CGCCACCTCT GACTTGAGG TCGATTTTG TGATGCTCGT CAGGGGGCG GAGCCTATGG AAAAACGCCA  
GCGGTGGAGA CTGAACCTCG AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTCGGATACC TTTTTCGGT

4551 GCAACGCGGC CTTTTTACGG TTCTGCGCT TTTGCTGCTC ATGTTCTTTC CTGCTGTATC  
CGTTGCGCCG GAAAAATGCC AAGGACCGG AAAACGAGTG TACAAGAAAG GACGCAATAG

4621 CCTTGATCT GTGGATAACC GTATTACCG CTTTGAAGTA GCTGATACCG CTCGCCGCG GCGAACGACC  
GGGACTAAGA CACCTATTGG CATAATGGG GAAACTCACT CGACTATGGC GAGCGGCGTC GGCCTGCTGG

4691 GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGC CAATACGCA ACCGCTTTC CCCGCGCGTT  
CTCGCGTCCG TCAGTCACTC GCTCCTTCG CTTCTCGCG GTTATGCGT TGGCGGAGG GGGCGCGCAA

4761 GGCCGATTCA TTAATGCAGC TGGCAGCA GGTTCCTCGA CTGGAAGCG GGCAGTGGC GCAACGCAAT  
CCGGCTAAGT AATTACGTG ACCGTGCTGT CCAAAGGCT GACCTTTCG CCGTCACTCG CGTTGCGTTA

4831 TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTGA CACTTTATGC TTCCGGCTCG TATGTTGTGT  
ATTACACTCA ATCGAGTGAG TAATCCGTG GGTCCGAAAT GTGAATATCG AAGGCCGAGC ATACAACACA

**FIG. 48H**

4901 GGAATTGTGA GCGGATAACA AATTACACACA GGAACACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA  
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTTGTGGA TACTGGTACT AATGCGGTTT CCGCGTTAAT

BssHII  
~ ~ ~ ~ ~

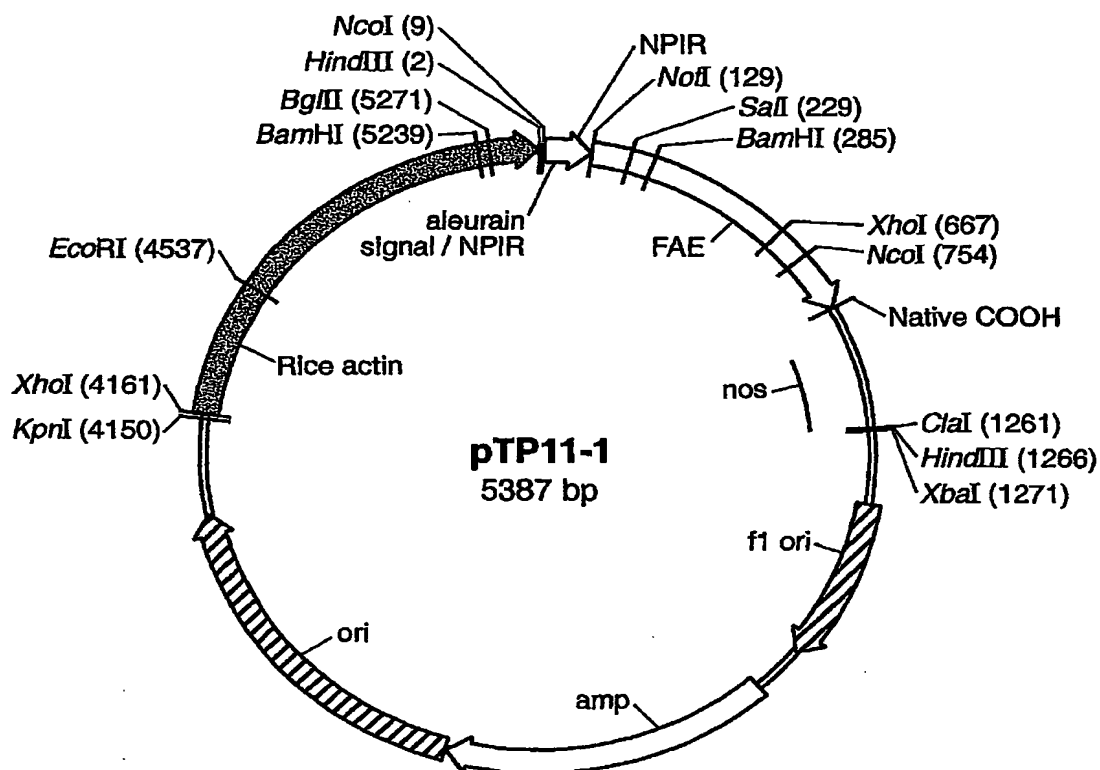
NcoI  
~ ~ ~ ~ ~

4971 ACCCTCCTA AAGGGAACAA AAGCTGGGTA C  
TGGGAGTGAT TTCCCTTGTT TCGACCCAT G

KpnI  
~ ~ ~ ~ ~

FIG. 48I



**FIG. 49A**

```

NGOI
~~~~~
HindIII
~~~~~
1  AAGCTTACCA TGGCCACGCG CCGGCTCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCGTCGCCG
    M A H A R V L L L A L A V L A T A A V A V
    NotI
    ~~~~~
 . A S S S F A D S N P I R P V T D R A A A S T .
71 TCGCCTCCTC CTCCTCCTC GCGACTCA ACCGATCCG GCGGTCCAC GACCGCGCG GCGCCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
 Sali
    ~~~~~
    AccI
    ~~~~~
211 D L C N I P S T I I K G E K I Y N S Q T D I N G
 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTTCACAA TTCTCAAACT GACATTACG
 BamHI
    ~~~~~
    . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AATAATCAC CGTCFTCCGT GGCACGTGTA GTGATACGAA
    . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAACTC GATACTAACT ACACCTTCAC GCCTTTTCGAC ACCCTACACG AATGCAACGG TTGTGAAGTA
    H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTTG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
    . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACAC CCTCGGCGCC TCCCTGGCGG CACTCAGTGC
    . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCCAGCTG TCTGGGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
    XhoI
    ~~~~~
631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
 CGCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA

```

FIG.\_49B

**FIG. 49C**

2451 TGAGTGATGA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT  
 2521 GCACAACATG GGGGATCATG TAACTCGCCT TGAATCGTTG GAACCGGAGC TGAATGAAGC CATACCAATC  
 2591 GACGAGCGTG ACACCAAGAT GCGTGTAGCA ATGGCAACAA ACTATTAAT ACTATTAAT GCGCAATC  
 2661 TTAATCTAGC TTCCCGGCAA CAATTAATAG ACTGATGGA GCGGATGGA GCGGATGGA GCGGATGGA  
 2731 CTCGGCCCTT CCGGCTGGCT GGTATATGTC TGATAAATCT GGAGCCGGTC AGCGTGGGTC TCGCGGTATC  
 2801 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCCTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA  
 2871 CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACCTGAT AAGCATTTGT AACTGTCTAGA  
 2941 CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAACCTT CATTTTAAAT TTAAGAAGAT CTAGGTGAAG  
 3011 ATCTTTTTCG AATATCTCAT GACCAAAATC CTTAAACGTG AGTTTTCGTT CCACCTGAGC TCAGACCCCG  
 3081 TAGAAAAGAT CAAAGGATCT TCTTGAATTC GTTTTTCCTT GCGGTATATC TCGTGTGTC AAACAATAAA  
 3151 ACCACCGCTA CCAGCGGTGG TTTTGTGTC GGTCTAGTGT AGCCGTAGTT AGGCCACCAT TTCAAGAACT  
 3221 TTCAGCAGAG CGCAGATACC AAATACTGTC CTCTAGTGT TAATCTGTT ACCAGTGGCT GCTGCGGCTG GCGATAAGTC  
 3291 CTGTAGCACC GCCTACATAC CTGCTCTGTC CAAGACGATA GTTACCCTGAT AAGCGCAGC CTTACAGCGT GAGCTATGAG  
 3361 GTGTCTTACC GGTGTGACT AGCCAGCTT GGAGCGAAG ACCTACACCG TCCGTTAAGC GGCAGGTCG GAACAGGAGA  
 3431 TCGTGCACAC AGCTCCCGAA GGGAGAAAG CGGACAGGTA TATAGTCTCTG TCGGTTCTCG CCACCTCTGA  
 3501 AAAGCGCCAC GCTTCCCGAA GGGAGAAAG CGGACAGGTA TATAGTCTCTG TCGGTTCTCG CCACCTCTGA  
 3571 GCGACAGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCTCTG TCGGTTCTCG CCACCTCTGA  
 3641 CTTGAGCGTC GATTTTGTG ATGCTCTGTC GGGGGGCGGA GCTATGGA AACGCCAGC AACCGGCTT  
 3711 TTTTACGGTT CTTGGCCCTT TGTGCTCACC GTTCTTCTCT CCGCTTCTCT CCGCTTCTCT CCGCTTCTCT  
 3781 GGATAACCGT ATTACCGCTT TTGAGTGAGC TGATACACCT TGATACACCT TGATACACCT CCGCTTCTCT  
 3851 TCAGTGAGCG AGGAAGCGGA AGAGCGCCA ATACGCAAC CCGCTTCTCT CCGCTTCTCT CCGCTTCTCT  
 3921 AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAGCGGG CAGTGAGGTC AATGAGGTTA AATGAGGTTA  
 3991 GCTCACTCAT TAGGCACCC AGGCTTTACA CTTTATGCTT CCGGCTGCTA CCGGCTGCTA CCGGCTGCTA  
 4061 GGATAACAAT TTCACACAGG AACAGCTAT GACCATGAT ACCTTGTATA ATTTGTATTA ATTTGTATTA  
 4131 GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTGAT TCATATGCTT GAGAAGAGAG TCGGGATAGT  
 4201 CCATAAATAA ACAAGGTAA GATTACCTGG TCATAAGTGA AACATCAGT TAAAGAGTGG TATAGATGTA  
 4271 ATATCGGTAA TAAAGGTGG CCCAAGTGA AATTACTCT TTTCTACTAT TATAAATTT GAGGATGTTT  
 4341 TGTGCGTACT TTGATACGTC ATTTTGTAT GAATGGGTTT TTAAGTTTAT TCGCGATTTG GAAATGCTA  
 4411 TCTGTATTG AGTCGGTTT TAAATTCGTT GCTTTTGTAA ATACAGAGG ATTTGTATTA ATTTGTATTA  
 4481 TAAAAAACC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATAT CAGGCGAAT CCACAATGAA  
 4551 CAATAATAAG ATTAATAAG CTTGCCCCCG TTGACGCGAT GGTATTTTT TCTAGTAAAA TAAAGATGTA  
 4621 ACTTAGACTC AAAACATTTA CAAAAACAAC CCTTAAAGTC CTTAAGCCCCA AAGTGTATG CAGGATCCAT

EcoRI

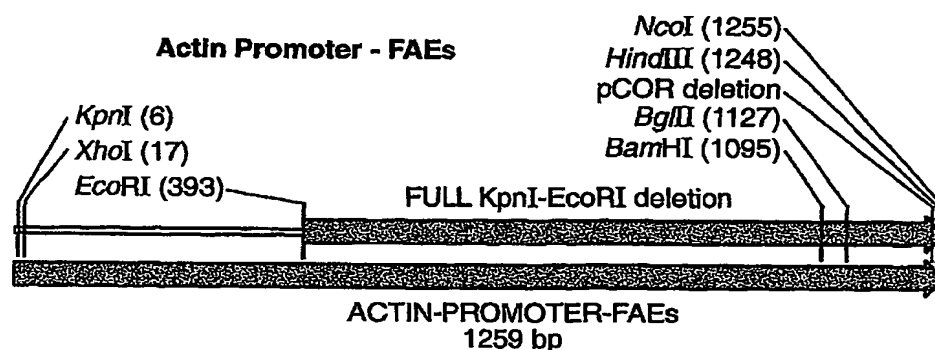
KpnI

XhoI

FIG. 49D

4691 AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC  
 4761 CCGGCACATAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA AAAAAAAA AGAAGAAAA  
 4831 AAAAGAAAAA GAAAAACAGC AGGTGGGTCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCCC  
 4901 CGACGAGGCC CGGCCCTCCC TCCCTCCCAAC CCFACCAACA CCACCAACAC CACCTCCTCC CCCCTCGCTG CCGGACGACG  
 4971 TCTCCTCCCA TCCCTCCCAAC CCFACCAACA CCACCAACAC CACCTCCTCC CCCCTCGCTG CCGGACGACG  
 5041 AGCTCCTTCCC CCTTCCCCCT CCGCCGCCGC CGGTAAACAC CCGCCCTCC TCTCTCTCTT TCTTCTCTTT  
 5111 TTTTCTTCTG CTGGGTCTCG ATCTTTGGCC TTGGTAGTTT GGTGGGCCGA GAGCGGCTTC GTGCCCCAGA  
 BamHI  
 5181 TCGGTGCGCG GGAGGGGCGG GATCTCGCG GATCTCGCTC CTGGCGTGAG TCGGCCCGGA TCCTCGCGGG  
 BglII  
 5251 GAATGGGGCT CTCGGATGTA GATCTCTTTT CTTTCTTCTT TTTGTGGTAG AATTGAATC CCTCAGCAT  
 5321 GTTCATCGGT AGTTTTTCTT TTCAATGATT GTGACAAATG CAGCCTCGTG CCGAGCTTTT TTGTAGC

FIG.-49E



|     | KpnI               |                    | XhoI               |                     |                   |                    |                    |  |
|-----|--------------------|--------------------|--------------------|---------------------|-------------------|--------------------|--------------------|--|
| 1   | <u>GGTACCGGGC</u>  | <u>CCCCCTCGA</u>   | <u>GGTCATTCAT</u>  | <u>ATGCTTGAGA</u>   | <u>AGAGAGTCGG</u> | <u>GATAGTCCAA</u>  | <u>AATAAAACAA</u>  |  |
|     | <u>CCATGGCCCG</u>  | <u>GGGGGGAGCT</u>  | <u>CCAGTAAGTA</u>  | <u>TACGAACTCT</u>   | <u>TCTCTCAGCC</u> | <u>CTATCAGGTT</u>  | <u>TTATTTTGTT</u>  |  |
| 71  | <u>AGGTAAGATT</u>  | <u>ACCTGGTCAA</u>  | <u>AAGTGAAAAC</u>  | <u>ATCAGTTAAA</u>   | <u>AGGTGGTATA</u> | <u>AGTAAAATAT</u>  | <u>CGGTAATAAA</u>  |  |
|     | <u>TCCATTCTAA</u>  | <u>TGGACCAGTT</u>  | <u>TTCAC TTTT</u>  | <u>TAGTCAATTT</u>   | <u>TCCACCATAT</u> | <u>TCATTTTATA</u>  | <u>GCCATTATTT</u>  |  |
| 141 | <u>AGGTGGCCCA</u>  | <u>AAGTGAAATT</u>  | <u>TACTCTTTTC</u>  | <u>TACTATTATA</u>   | <u>AAAATTGAGG</u> | <u>ATGTTTGTCT</u>  | <u>GGTACTTTGA</u>  |  |
|     | <u>TCCACCGGGT</u>  | <u>TTCACTTTAA</u>  | <u>ATGAGAAAAG</u>  | <u>ATGATAATAT</u>   | <u>TTTTAACTCC</u> | <u>TACAAAACAG</u>  | <u>CCATGAAACT</u>  |  |
| 211 | <u>TACGTCA TTT</u> | <u>TTGTATGAAT</u>  | <u>TGGTTTTTAA</u>  | <u>GTTTATTTCG</u>   | <u>GATTTGGAAG</u> | <u>TGCATATCTG</u>  | <u>TATTTGAGTC</u>  |  |
|     | <u>ATGCAGTAAA</u>  | <u>AACATACTTA</u>  | <u>ACCAAAAATT</u>  | <u>CAAATAAGCG</u>   | <u>CTAAACCTTT</u> | <u>ACGTATAGAC</u>  | <u>ATAAACTCAG</u>  |  |
| 281 | <u>GGTTTTTAAG</u>  | <u>TTCGTTGCTT</u>  | <u>TTGTAAATAC</u>  | <u>AGAGGGATT</u>    | <u>GTATAAGAAA</u> | <u>TATCTTTAAA</u>  | <u>AAACCCATAT</u>  |  |
|     | <u>CCAAAAATTC</u>  | <u>AAGCAACGAA</u>  | <u>AACATTTATG</u>  | <u>TCTCCCTAAA</u>   | <u>CATATTCTTT</u> | <u>ATAGAAATTT</u>  | <u>TTTGGGTATA</u>  |  |
|     |                    |                    |                    |                     | <u>EcoRI</u>      |                    |                    |  |
| 351 | <u>GCTAATT TGA</u> | <u>CATAATTTT</u>   | <u>GAGAAAAATA</u>  | <u>TATATT CAG</u>   | <u>GGAATTCAC</u>  | <u>AATGAACAAT</u>  | <u>AATAAGATTA</u>  |  |
|     | <u>CGATTAAACT</u>  | <u>G TATTAAAA</u>  | <u>CTCTTTT TAT</u> | <u>ATATAAGTCC</u>   | <u>GCTTAAGGTG</u> | <u>TTACTTGTTA</u>  | <u>TTATTCTAAT</u>  |  |
| 421 | <u>AAATAGCTTG</u>  | <u>CCCCCGTGC</u>   | <u>AGCGATGGGT</u>  | <u>ATTFTTCTA</u>    | <u>GTAAAATAAA</u> | <u>AGATAAACTT</u>  | <u>AGACTCAAAA</u>  |  |
|     | <u>TTTATCGAAC</u>  | <u>GGGGGCAACG</u>  | <u>TCGCTACCCA</u>  | <u>TAAAAAAGAT</u>   | <u>CATTTTATTT</u> | <u>TCTATTG TAA</u> | <u>TCTGAGTTTT</u>  |  |
| 491 | <u>CATTTACAAA</u>  | <u>AACAACCCCT</u>  | <u>AAAGTCCTAA</u>  | <u>AGCCCAAAGT</u>   | <u>GCTATGCACG</u> | <u>ATCCATAGCA</u>  | <u>AGCCCAGCCC</u>  |  |
|     | <u>GTAAATGTTT</u>  | <u>TTGTTGGGGA</u>  | <u>TTTCAGGATT</u>  | <u>TCGGGT T TCA</u> | <u>CGATACGTGC</u> | <u>TAGGTATCGT</u>  | <u>TCGGGTCGGG</u>  |  |
| 561 | <u>AACCCAACCC</u>  | <u>AACCCAACCC</u>  | <u>ACCCAGTGC</u>   | <u>AGCCAAC TGG</u>  | <u>CAATAGTCT</u>  | <u>CCACCCCGG</u>   | <u>CAC TATCACC</u> |  |
|     | <u>TTGGGT TGGG</u> | <u>TTGGGT TGGG</u> | <u>TGGGGT CAG</u>  | <u>TCGGTTGACC</u>   | <u>GTTTATCAGA</u> | <u>GGTGGGGGCC</u>  | <u>GTGATAGTGG</u>  |  |
| 631 | <u>GTGAGTTGTC</u>  | <u>CGCACCACCG</u>  | <u>CACGTCTCGC</u>  | <u>AGCCAAAAAA</u>   | <u>AAAAAAAGAA</u> | <u>AGAAAAAAA</u>   | <u>GAAAAAGAAA</u>  |  |
|     | <u>CACTCAACAG</u>  | <u>GCGTGGTGGC</u>  | <u>GTGCAGAGCG</u>  | <u>TCGGTTTTTT</u>   | <u>TTTTTTTCTT</u> | <u>TCTTTTTTTT</u>  | <u>CTTTTCTTTT</u>  |  |
| 701 | <u>AACAGCAGGT</u>  | <u>GGGTCCGGGT</u>  | <u>CGTGGGGGCC</u>  | <u>GGAAAAGCGA</u>   | <u>GGAGGATCGC</u> | <u>GAGCAGCGAC</u>  | <u>GAGGCCCGGC</u>  |  |
|     | <u>TTGTCGTCCA</u>  | <u>CCCAGGCCCA</u>  | <u>GCACCCCGCG</u>  | <u>CCTTTTCGCT</u>   | <u>CCTCCTAGCG</u> | <u>CTCGTCGCTG</u>  | <u>CTCCGGGCGC</u>  |  |

**FIG. 50A**

771 CCTCCCTCCG CTTC~~CAAAGA~~ AACGCCCCC ATCGCCACTA TATACATACC CCCCCTCTC CTCCCATCCC  
GGAGGGAGGC GAAGGTTTCT TTGCGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG

841 CCCAACCCTA CCACCACCAC CACCACCACC TCCTCCCCC TCGCTGCCGG ACGACGAGCT CCTCCCCTCT  
GGGTTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCTCCGC CGCCGCCGGT AACCACCCCG CCCCTCTCCT CTTCTTTTCT CCGTTTTTTT TTCTGCTCTG  
GGGGGAGGCG GCGGCGGCCA TTGGTGGGCG GGGGAGAGGA GAAACAAAGA GGCAAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCG CCCAGATCGG TCGCGGGGAG  
CAGAGCTAGA AACCGGAACC ATCAAACCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

BamHI

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCGGATCCT CGCGGGGAAT GGGGCTCTCG  
CCCGCCCTAG AGCGCCGACC GCAGAGGCC GCACCTAGCC GGGCCTAGGA GCGCCCTTA CCGGAGAGC

BglII

1121 GATGTAGATC TTCTTTCTTT CTCTTTTTTG TGGTAGAATT TGAATCCCTC AGCATTGTTT ATCGGTAGTT  
CTACATCTAG AAGAAAGAAA GAAGAAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII NcoI

1191 TTTCTTTTCA TGATTTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTTTGT AGGTAGAAGC TTACCATGG  
AAAGAAAAGT ACTAAACACT GTTTACGTCG GAGCACGCCT CGAAAAACA TCCATCTTCG AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9,  
pJQ3.2 and pJO6.3.

**FIG.\_50B**

**ALEURAIN\_deleted NPIR (Apoplast) Structure and Sequence**

**ALEURAIN-NPIR-DEL**  
93 bp

+1            M   A   H   A   R   V   L   L   L   A   L   A   V   L   A   T   A   A   V   A  
HindIII NcoI

\*\*\*\*\*

1   AAGCTTACCA TGGCCACGC CCGCTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTGCGCG  
TTCGAATGGT ACCGGGTGCG GCGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC

+1 V   A   S   S   R   A   A

NotI

\*\*\*\*\*

71   TCGCCTCCTC CCGCGCGGCC GCC  
AGCGGAGGAG GCGCGCGCGG CGG

**FIG.\_51**



**SEE1 ( Senescence enhanced ) PROMOTER sequence**

```

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
 PstI XbaI XbaI
      ~~~~~
71  AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
      SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
 SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTTG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
      PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG

```

**FIG. 52****SEE1 ( Senescence enhanced ) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence**

```

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
 PstI XbaI XbaI
      ~~~~~
71  AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
      SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
 SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTTG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
      PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
 M A H G R I L F L A L A V L
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
 BssHII
      ~~~~~
      NotI
      ~~~~~
911 . A T A A V A A A S L A D S N P I R P V T E R A .
 TGGCCACCGC CGCGGTGGCC GCCGCATCCT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
 NotI
      ~~~~~
981 . A A
      GGCCGCC

```

**FIG. 53**

## INTERNATIONAL SEARCH REPORT

Intern Application No  
PCT/US 01/43588

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 C12N15/82

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, EPO-Internal, PAJ, BIOSIS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                   | Relevant to claim No. |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X          | US 6 143 543 A (MICHELSEN BIRGIT ET AL)<br>7 November 2000 (2000-11-07)<br>see the whole document                                                                                                                                                                                                                                                                    | 1-3,<br>34-36, 74     |
| Y          |                                                                                                                                                                                                                                                                                                                                                                      | 4-33,<br>37-73        |
| Y          | DE VRIES R P ET AL: "The faeA genes from<br>Aspergillus niger and Aspergillus<br>tubingensis encode ferulic acid esterases<br>involved in degradation of complex cell<br>wall polysaccharides."<br>APPLIED AND ENVIRONMENTAL MICROBIOLOGY,<br>vol. 63, no. 12, December 1997 (1997-12),<br>pages 4638-4644, XP002203731<br>ISSN: 0099-2240<br>see the whole document | 4-33,<br>37-73        |
|            | -/-                                                                                                                                                                                                                                                                                                                                                                  |                       |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

27 June 2002

Date of mailing of the international search report

12/07/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Grosskopf, R

# INTERNATIONAL SEARCH REPORT

Inten ☐ Application No  
PCT/US 01/43588

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                         |                       |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category *                                           | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                      | Relevant to claim No. |
| A                                                    | <p>GARCIA-CONESA MARIA-TERESA ET AL: "A cinnamoyl esterase from <i>Aspergillus niger</i> can break plant cell wall cross-links without release of free diferulic acids." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 266, no. 2, December 1999 (1999-12), pages 644-652, XP002203732<br/>ISSN: 0014-2956</p> |                       |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern

Application No

PCT/US 01/43588

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s)  | Publication<br>date |
|-------------------------------------------|---------------------|-----------------------------|---------------------|
| US 6143543                                | A                   | 07-11-2000 GB 2301103 A , B | 27-11-1996          |

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**